A Bayesian Approach for Adaptively Modulated Signals Recognition in Next-Generation Communications

Bin Li, Shenghong Li, Jia Hou, Junqiang Fu, Chenglin Zhao, and Arumugam Nallanathan, *Senior Member, IEEE*

***Abstract*—By promoting spectrum efficiency and transmission reliability, link adaptation (e.g., adaptive modulation) is one of the enabling technologies for next-generation 5G communications. In this paper, we investigate the recognition of adaptively modulated signals, which remains still as an unexploited area as far as we are aware, especially in the presence of time-varying fading channels. A unified model, relying on the dynamic state-space approach, is formulated, which thoroughly characterizes the coupling relationship between two hidden states, i.e., unknown modulation schemes and fading channels. In contrast to existing schemes marginalizing directly out random fading effects, a joint estimation paradigm, which relies on the Bayesian stochastic inference and a maximum *a posteriori* criterion, is developed to acquire time-correlated fading states sequentially, at the same time of recognizing unknown modulation schemes. In order to alleviate the computation complexity, two simplified schemes, i.e., *fading-driven* and *goal-oriented*, are designed. It is demonstrated that, by fully exploiting the underlying dynamics of the estimated fading gain which is modeled by a discrete-states Markov chain, the recognition performance of adaptive modulations will be improved significantly. The proposed system model and a sequential estimation framework, by providing additionally the dynamic fading channels, may be of great promise to more flexible and effective link adaptations.**

***Index Terms*—Next-generation communications, 5G, link adaptation, adaptive modulation, dynamic modulation recognition, time-varying fading, sequential Bayesian estimation.**

I. Introduction

ADAPTIVE modulation is one of the key enabling techniques for next-generation communications (e.g., 802.11ac and LTE-A) [1]–[3], which is intended to promote the link adaptation performance. In order to maximize both the spectrum efficiency and transmission reliability, the optimum modulation scheme should accommodate the current channel condition [4], e.g., the probed channel state information (CSI) [5]. As a flip-over procedure, the modulation recognitions, aiming to identify unknown modulated signals from a prescribed finite set, should be carried out by receivers premised on a sequence of noisy observations [6]. Despite intensive studies in the military community [7]–[9], modulation recognitions have drawn wide interests recently in commercial communications, e.g., dynamic spectrum sharing [10] and interference awareness (e.g., device-to-device) applications [11].

As in most information interception scenarios, e.g., non-cooperative detection or interference mitigation, the modulation recognition has to be accomplished in a *blind* manner [6]. That means, there is no *a priori* information (i.e., the pilot or preamble sequence) can be utilized. Typical blind recognitions may occur in either homogeneous or heterogeneous communications, in which the preamble (or overhead) is usually unavailable due to the consideration of transmission efficiency or the lack of coordination [10], [11]. Taking a homogeneous cellular network for example, multiple-users (e.g., with multiple input and multiple output) may be not perfectly orthogonal due to imperfect CSI. As a consequence, the multi-user interference will be aroused, where the modulation mode usually is unknown to other receivers [11]. For heterogeneous networks operating in the overlapped frequency band, e.g., wireless local area networks (WLANs) and ultra-wideband (UWB), the modulation scheme of the interfering signals from another uncoordinated system (e.g., 802.11ac) may remain also blind to the interfered devices (e.g., UWB).

Traditional techniques on modulation recognitions may be summarized into two categories, i.e., feature-based methods and likelihood-based approaches [6]. The first scheme involves two steps, i.e., feature selections (e.g., cumulates, wavelets or correlations) and pattern classifications (e.g., blind clustering, neural networks or support vector machine) [12]–[17], which is practically immune to imperfect conditions, e.g., model mismatch or completely blind cases [18]. The second one, instead, exploits the statistical likelihood of received signals (e.g., the channel information and noise variance) [18]–[24], which is optimal under a common assumption of the *uniform* prior, in the sense that it minimizes the average probability of false classifications of a finite number of candidates. Such maximum likelihood (ML) methods, therefore, are of great promise to realistic applications

Manuscript received January 05, 2015; revised April 28, 2015; accepted May 12, 2015. Date of publication June 02, 2015; date of current version July 09, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Subhrakanti Dey. This work was supported by Natural Science Foundation of China (NSFC) under Grants 61471061 and the Fundamental Research Funds for the Central Universities under Grant 2014RC0101.

B. Li, J. Fu, and C. Zhao are with the School of Information and Communication Engineering (SICE), Beijing University of Posts and Telecommunications (BUPT), Beijing, 100876, China (e-mail: stonebupt@gmail.com).

S. Li is with Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: shli@stju.edu.cn).

J. Hou is with School of Electronic and Information Engineering, Soochow University, Suzhou 215006, China (e-mail: houjia@suda.edu.cn).

A. Nallanathan is with the Department of Informatics Engineering, King's College London, London, WC2R2LS, U.K. (e-mail: nallanathan@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2015.2440189

and has been studied extensively [6]. To further make this computationally complex scheme viable, simplified or near-optimal ML schemes with the lower complexity burden have been investigated, for example, the look-up tables (LUTs) [10] and the mixed online/offline ML techniques [18].

Other than the implementation complexity, another major obstacle in applying the ML-based recognition technique is that it relies heavily on the complete CSI (e.g., the fading channel). In order to cope with unknown channel parameters, the average likelihood ratio test (ALRT) and a family of similar methods have been proposed [6]. In [22] and [25], a joint estimation scheme is investigated based on the method of moments (MOM), with which the fading chanel is derived from several statistical moment equations. In order to precisely calculate higher-order moments [22], [25], the computation burden and the time consumption of this approach may be unaffordable.

As far as future mobile scenarios are concerned [26], adaptive modulations would be implemented in accordance with variant fading channel [27]. Adaptive recognitions under time-varying fading effects, furthermore, will bring two new challenges. First, the modulation schemes would be selected adaptively according to time-correlated fading conditions, which will tend to be mutually dependent. As a consequence, the modulation candidates, which are coupled with time-varying channels, are no longer uniformly distributed, i.e., with the *a priori* density changing with times. In such cases, most ML-based approaches become not optimal. Second, the varying unknown fading will arouse the noticeable *uncertainty* in received signals, making the likelihood function not accurate anymore. More importantly, the next-generation communications demand for more flexible and effective link adaptations, indicating the CSI (i.e., fading channels) should also be acquired sequentially in real time. Most conventional schemes, e.g., the MOM method, fail unfortunately to track time-varying fading channel adaptively and may be less attractive in realistic applications.

To the best of our knowledge, there are few works in the literature reported on the design of *dynamic* or *adaptive* modulation recognition in more challenging time-varying fading channels. In this paper, we present a promising Bayesian recognition approach for such adverse environments. A major innovation of our scheme is that, relying on a new formulated dynamic system model and a sequential estimation algorithm, the dynamically transitional property of time-correlated channels will be fully utilized. Adaptively modulated signals, therefore, can be recognized by estimating time-varying fading gains jointly. To sum up, the main contributions of this work are summarized as follows.

1) A new mathematical formulation, which is premised on the dynamic state-space model (DSM), is proposed. In sharp contrast to existing recognition models, the dynamic behaviors of realistic time-varying fading channels, which are characterized by a discrete-state Markov chain (DSMC), is concerned. Meanwhile, the coupling relationship between time-dependent unknown fading gains and adaptively modulated signals is taken fully into accounts.
2) A sequential Bayesian estimation algorithm, which derives time variant fading channels and unknown modulation candidates jointly, is developed. By maximizing the joint *a posteriori* density, the fading gains accompanying unknown modulation candidates are estimated sequentially. Given different maps from the signal to noise ratio (SNR) to modulation candidates, two simplified schemes are developed in order to alleviate the computational complexity. With the maximum *a posteriori* (MAP) paradigm, the underlying dynamics of time-correlated fading channel could be thoroughly exploited and, consequently, the recognition performance is enhanced.
3) The estimation performances of both fading channels and the average recognition accuracy are investigated based on numerical experiments. It is shown that, even with time-varying fading propagation, the adaptive modulation recognition will be effectively realized, by acquiring unknown fading gains simultaneously. Compared to the existing counterpart, e.g., the ALRT scheme, the proposed scheme may promote the recognition accuracy significantly, by exploiting the dynamics of time-correlated fading gains.
4) The proposed scheme creates a unified paradigm for likelihood-based adaptive modulation recognitions. The new algorithm can be considered as a *dynamic* MAP approach, as distinguished from traditional *static* likelihood, which provides an effective tool to handle various modulation recognitions. Despite focusing on unknown fading gains, the formulated DSM as well as the joint Bayesian estimation algorithm can be extended conveniently to some other realistic non-ideal scenarios, e.g., in the case of unknown noise variance or timing deviations. In general, the proposed scheme may lead to the more flexible and effective link adaptation solutions for next-generation 5G communications.

The rest of the article is structured as following. In Section II, a new general DSM is formulated for adaptive modulation recognition in realistic time-varying fading channels. Then, a sequential Bayesian algorithm is introduced in Section IV. On this basis, two simplified recursive algorithms, which jointly estimate dynamic fading channels and unknown modulation candidates, are further designed. In Section V, simulations and performance analysis are provided. Finally, we conclude this investigation in Section VI.

II. System Model and Problem Formulation

A. Link Adaptation With Time-Varying Fading

An illustration of adaptive modulation in the context of time-correlated fading channel is given by Fig. 1. As shown in the *right* sub-figure, the fading gain α_n is assumed as common to follow the *a priori* Rayleigh distribution, i.e.,

$$f(\alpha) = \frac{\alpha}{\sigma^2} \times \exp\left(-\frac{\alpha^2}{2\sigma^2}\right), \tag{1}$$

where σ^2 denotes the variance of Rayleigh distribution.

For the emerging applications with unrestrained mobile users [1], [26], the fading channel α_n may become time variant [27], as shown by the *middle* figure. For the common slowly varying channel, α_n of discrete time n will be related with the past fading state α_{n-1}. As in the *left* sub-figure, the modulation candidate is selected according to the current channel condition.

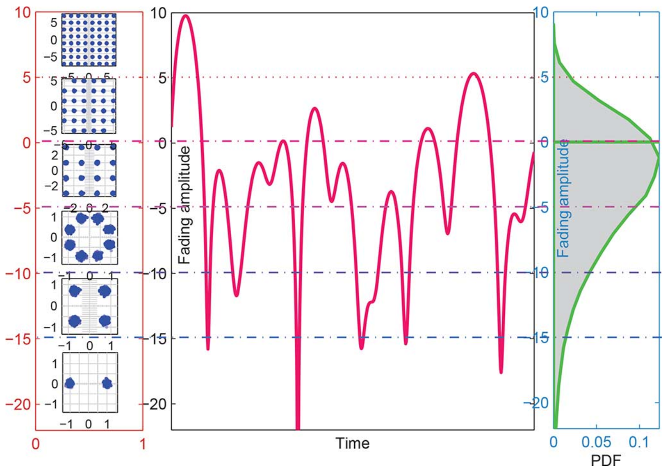


Fig. 1. Adaptive modulations in time-varying flat-fading channels. The *left* gives multiple modulation candidates. The *middle* figure shows Rayleigh fading amplitudes varying with time. The *right* demonstrates the statistical probability density function (PDF) of fading amplitudes.

B. Dynamic State-Space Model

In the consideration of more challenging dynamic modulation recognition under time-varying fading effects, a new stochastic DSM is formulated as follows.

$$\alpha_n = A(\alpha_{n-1}), \tag{2}$$

$$c_n = T(\alpha_n), \tag{3}$$

$$\mathbf{z}_n = Z(\alpha_n, \theta_n, c_n, w_{n,m}). \tag{4}$$

Here, (2)–(3) are referred to as *dynamic equations*, while (4) is the *measurement equation* [28].

1) The stochastic transitional function $A(\cdot) : \mathbb{R}^1 \mapsto \mathbb{R}^1$ gives the dynamic behaviors of hidden time-varying fading states $\alpha_n \in \mathcal{A} \subset \mathbb{R}^1$ of the $(n-1)$th time slot, which is characterized by the 1st-order Markov process.
2) The other transitional function $T(\cdot) : \mathbb{R}^1 \mapsto \mathbb{Z}^1$ prescribes the map from the current fading state α_n to a single modulation candidate $c_n \in \mathcal{C} \subset \mathbb{Z}^1$.
3) The observation function $Z(\cdot) : \mathbb{R}^M \mapsto \mathbb{R}^M$ describes the coupling relationship between the hidden states (i.e., α_n, c_n and the channel phase θ_n) and the observation $\mathbf{z}_n \subseteq \mathbb{R}^M$, which is composed of M discrete samples $z_{n,m}(m = 0, 1, \cdots, M-1)$.

Note that, based on two considerations, the hidden state α_n in (2) accounts only for the time-varying fading gain. First, the temporal channel condition, incarnated essentially by the SNRs (or link gains), is the dominant factor of adaptively selecting a modulation scheme. Second, the *time-dependence* of realistic fading is mainly exercised by its gains, while the unknown channel phase θ_n is usually distributed uniformly [29]. In subsequent analysis, therefore, the channel phases θ_n will be treated separately as another unknown quantity.

For the ease of analysis, two points are assumed in the established DSM model.

1) The random noise involved by the measurement process, denoted by $w_{n,m} \in \mathbb{R}^1$, is assumed to be the zero-mean additive white Gaussian noise (AWGN), i.e., $w_{n,m} \sim \mathcal{CN}(0, \sigma_w^2)$, which is independent and identically distributed (i.i.d) and remains also independent of two hidden states α_n and c_n.
2) The fading channel is assumed to be slowly varying. The fading gain α_n will be invariant in L_a transmission slots, where L_a is inversely proportional to the maximum Doppler frequency shift f_D, i.e., $L_a \propto 1/f_D$. That means, α_n will be independent of different sample index m.

Next, we will elaborate on each dynamic equations and the measurement equation.

C. Time-Varying Fading Channel

For emerging applications with mobile devices (e.g., LTE-A) or relative movements, the time-correlated fading channel will be commonly encountered. The dynamic behavior of time-dependent fading channels, which evolves with times, can be modeled as a DSMC [29]. It is shown that the DSMC will reflect the dynamic nature of time-varying fading channels effectively [30], [31], which is also a sufficiently good match to the statistical model, e.g., Clarke's model.

For most realistic fading, the stochastic property of its time correlations is stationary or ergodic, e.g., characterized by a wide sense stationary uncorrelated scattering (WSSUS) model. Thus, the DSMC model is assumed to be stationary, which is reasonable in many communication scenarios, e.g., with the channel statistics changing slowly over time [32]–[34], [38]. The transitional probabilities of fading channels, as a consequence, will be independent of n.

For convenience, an indecomposable DSMC is concerned. That is, if we denote the stationary probability vector $\boldsymbol{\pi} = [\pi_0, \pi_1, \cdots, \pi_{K-1}]^T$ with $\pi_k \triangleq \Pr(\alpha_n = \mathsf{A}_k)$, we will have $\mathbf{P}^T\boldsymbol{\pi} = \boldsymbol{\pi}$ [34], where $\mathsf{A}_k \in \mathcal{A}$ is the representative fading state and $\mathbf{P}_{K\times K} = \{P_{k_1\to k_2}, k_1, k_2 \in [0, 1, \cdots, K-1]\}$ denotes the transitional probability matrix (TPM) of α_n, i.e.,

$$\mathbf{P}_{K\times K} = \begin{bmatrix} P_{0\to 0} & P_{0\to 1} & \cdots & P_{0\to(K-1)} \\ P_{1\to 0} & P_{1\to 1} & \cdots & P_{1\to(K-1)} \\ \vdots & \vdots & \ddots & \vdots \\ P_{(K-1)\to 0} & P_{(K-1)\to 1} & \cdots & P_{(K-1)\to(K-1)} \end{bmatrix}, \tag{5}$$

where each elements $P_{k_1\to k_2}$ specifies the transitional probability from the state k_1 to the state k_2, i.e.,

$$P_{k_1\to k_2} \triangleq \Pr\left(\alpha_{n'} = \mathsf{A}_{k_2} | \alpha_{n'-1} = \mathsf{A}_{k_1}\right). \tag{6}$$

In the above formulation, the discrete state $\alpha_{n'}$ is treated as an unknown output belonging to one feasible set $\mathcal{A} = \{\mathsf{A}_k, k \in [0, 1, \cdots, K-1]\}$. Here, $n' = \lfloor n/L_a \rfloor$ refers in particular to the time slot where the transitions of slowly varying α_n occurs. In order to determine a specific DSMC, e.g., $\mathcal{A}$ and $\mathbf{P}_{K\times K}$, firstly the nonnegative fading amplitude α (or the received SNR) will be partitioned to K non-overlapping regions, relying on a group of SNR thresholds $\mathcal{S} = \{s_1, s_2, \cdots, s_K\}$ or, equivalently, another group of gain thresholds $\mathcal{G} = \{g_1, g_2, \cdots, g_K\}$. If we further define $g_0 \triangleq 0$ and $g_K \triangleq \infty$, then we have:

$$\mathcal{G} = \{[g_0, g_1), [g_1, g_2), \ldots, [g_{K-1}, g_K)\}. \tag{7}$$

As a feasible assumption, here the SNR (or fading gain) thresholds of the adaptive modulation scheme protocol (e.g., IEEE 802.11ac) is known as *a priori* knowledge.

Given the statistic PDF of fading gain, e.g., the aforementioned Rayleigh PDF $f(\alpha)$, the steady probability that α_n resides in the kth region $[g_k, g_{k+1})$, and the corresponding discrete state are derived respectively by:

$$\pi_k = \int_{g_k}^{g_{k+1}} f(\alpha)d\alpha, \mathsf{A}_k = \frac{1}{\pi_k} \times \int_{g_k}^{g_{k+1}} \alpha f(\alpha)d\alpha. \quad (8)$$

Under an equiprobable partition, i.e., $\pi_k \triangleq 1/K$, we may easily derive the partitioning bounds by $g_k^2 = -2\sigma^2 \ln(1-k/K)$ [29], [34]. Finally, the transitional probability $P_{k_1 \to k_2}$ will be determined from:

$$\begin{aligned} P_{k_1 \to k_2} &= \Pr\left\{\alpha_{n'} \in [g_{k_2}, g_{k_2+1}) | \alpha_{n'-1} \in [g_{k_1}, g_{k_1+1})\right\} \\ &= \frac{1}{\pi_{k_1}} \int_{g_{k_2}}^{g_{k_2+1}} \int_{g_{k_1}}^{g_{k_1+1}} f(\alpha_{n'-1}, \alpha_{n'}) d\alpha_{n'-1} d\alpha_{n'}, \end{aligned} \quad (9)$$

where $f(\alpha_{n'-1}, \alpha_{n'})$ is the *bi-variate* Rayleigh joint PDF.

For analytical simplicity, here we adopt the 1st-order FSMC model [33], which is usually viewed as a good match with the popular statistical fading models (e.g., Clarke's model). Thus, the current fading state α_n is only associated with the previous state α_{n-1}, while statistically independent of all remaining past states α_{n-n_0} and future states α_{n+n_0} $(n_0 \geq 1)$. To be specific, we now have $P_{k_1 \to k_2} = 0$ for $|k_1 - k_2| > 1$. So, the TPM $\mathbf{P}_{K\times K}$ can be further simplified to (10), shown at the bottom of the page.

In practice, we evaluate the transitional probabilities by utilizing the level crossing rate (LCR) N_k, which refers to the number of times per second that the fading amplitude crosses g_k in a downward direction [29]. So, we have:

$$N_k = \frac{\sqrt{2\pi} f_D g_k}{\sigma} \times \exp\left(-\frac{g_k^2}{\sigma^2}\right), \quad (11)$$

Relying on LCR, the transitional probability will be approximated by $P_{k_1 \to k_2} \simeq N_{k_2}/R_{k_1}$ [32]–[34]. Here, $R_{k_1} = \pi_{k_1}/(MT_s)$ is the average number of sensing slots per second in the state k_1, and T_s denotes the sample time.

D. Mapping Criterion

During the transmission slot n, a modulation scheme $c_n = C_l \in \mathcal{C} \subset \mathbb{Z}^1$ will be adaptively selected by a transmitter, coinciding with the dynamic channel condition or the fading region $\mathsf{A}_k \in \mathcal{A} \subset \mathbb{R}^1, k = \{1, 2, \cdots, K\}$. The modulation cardinality is denoted by $|\mathcal{C}| = L$. In the receiver-end, the blind recognition[1] will be performed and, as distinguished from existing schemes, the fading gain (or region) will be estimated jointly.

[1]Note that, the term "blind" refers specifically to the absence of *a priori* preamble sequence. Strictly speaking, the designed joint estimation scheme should rely on some partial information on channels, e.g., the *a priori* fading distribution and its variance. In order to differentiate from the pilot (or training preamble) assisted recognition schemes, however we may still refer the designed semi-blind algorithm as a blind one.

Then, the modulation candidate will be derived according to the specific mapping criterion in (3), and the estimated fading condition (i.e., channel gain) is reported to the transmitter via a feedback channel.

We firstly consider a simple single-to-single (S2S) map $T_{S2S}(\cdot) : \mathbb{R}^1 \to \mathbb{Z}^1$, which coordinates each fading gain $\alpha_n \in \mathcal{A}$ with a unique modulation $c_n = C_l$. In the case, we have $L = K$, i.e.,

$$T_{S2S}(\alpha_n = \mathsf{A}_k) = C_l, l = k. \quad (12)$$

In the other many-to-single (M2S) case, instead, there may exist K_l fading states $(2 \leq K_l \leq K)$ mapped to one single modulation C_l, i.e.,

$$T_{M2S}(\alpha_n = \mathsf{A}_k \in \mathcal{A}_l) = C_l, |\mathcal{A}_l| = K_l, \quad (13)$$

where $\mathcal{A}_l \subset \mathcal{A}$ denotes a *sub*-set of fading states having the same assignment with $c_n = C_l$ $(0 \leq l < |\mathcal{C}|)$, i.e.,

$$\mathcal{A}_l \triangleq \{\mathsf{A}_k | T_{M2S}(\mathsf{A}_k) = T_{M2S}(\mathsf{A}_{k'}) = C_l, |k - k'| \leq K_l - 1\}.$$

Note that, despite a unique modulation candidate of the M2S map, in practice different coding rates (e.g., the convolutional code of 1/2 rate and 2/3 rate) will be used among $\mathcal{A}_l$, which, nevertheless, will not be resolved by a modulation recognizer.

Based on the elaboration above, it is found that the *a priori* density of adaptive modulations now becomes time-correlated and *not* uniformly distributed. More specifically, in the nth slot the prior probability of the lth modulation candidate, denoted by $p(c_n = C_l)$, is given by:

$$p(c_n) = \begin{cases} p_{n|n-1}(c_n = C_l | c_{n-1} = C_{l'}), & |l - l'| \leq l_0, \quad (14a) \\ 0, & |l - l'| > l_0, \quad (14b) \end{cases}$$

where $p_{n|n-1}(c_n = C_l | c_{n-1} = C_{l'})$ accounts for the time-dependent prior transitional probabilities of the nth transmission slot, e.g., from the l'th candidate to the lth one. Here, l_0 denotes the maximum transitional interval between adjacent candidates. For the concerned 1st-order DSMC channel, we will have $l_0 \leq 1$ for both the S2S and M2S maps.

Due to the coupling relationship of $T(\cdot)$, the prior transitional density will be further associated with the dynamic transitions of fading states. Taking the S2S map for example, i.e., $L = K$ and $T(\mathsf{A}_{k_1}) \neq T(\mathsf{A}_{k_2})$ when $k_1 \neq k_2$, the prior transitional probability is specified by:

$$p_{n|n-1}(c_n = C_l | c_{n-1} = C_{l'}) = p_{n|n-1}(\alpha_n = \mathsf{A}_{k_2} | \alpha_{n-1} = \mathsf{A}_{k_1}). \quad (15)$$

$$\mathbf{P}_{K\times K} = \begin{bmatrix} P_{0\to 0} & P_{0\to 1} & 0 & 0 & \cdots & 0 & 0 & 0 \\ P_{1\to 0} & P_{1\to 1} & P_{1\to 2} & 0\cdots & 0 & 0 & 0 & \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & P_{(K-2)\to(K-3)} & P_{(K-2)\to(K-2)} & P_{(K-2)\to(K-1)} \\ 0 & 0 & 0 & 0 & \cdots & 0 & P_{(K-1)\to(K-2)} & P_{(K-1)\to(K-1)} \end{bmatrix}. \quad (10)$$

For the other M2S map, i.e., $L < K$, we have:

$$p_{n|n-1}(c_n = C_l|c_{n-1} = C_{l'}) = \sum_{\mathsf{A}_k \in T^{-1}(c_n=C_l)} p_{n|n-1}(\alpha_n = \mathsf{A}_k \in \mathcal{A}_l|\alpha_{n-1} = \mathsf{A}_{k'} \in \mathcal{A}_{l'}), \tag{16}$$

where $T^{-1}(c_n = C_l)$ results in a sub-set $\mathcal{A}_l \subset \mathcal{A}$ of the fading states.

It is noted that, in contrast to more complicated bit-loading mechanisms allocating various bit/modulation modes to different sub-carriers [35], in the threshold-based mapping scheme a single modulation is adopted by the transmitter as in most recognition scenarios [3], [6], which, however, requires also the side channel condition (i.e., the estimated fading states or the SNR regions).

From the above analysis, the formulated adaptive recognition keeps dramatically different from existing models in which both the *dynamic* behaviors of fading channels and the *non-uniform* prior densities of modulation candidates are rarely concerned. Such static recognition schemes, therefore, will be less attractive in future adaptive modulation scenario.

E. Observations

With unknown time-varying fading channels and the blind modulation scheme, in practice the received equivalent signal is written to:

$$z_n(t) = \alpha_n e^{-j\theta_n} \times \sum_{m=1}^{\infty} \sqrt{E_s}\mathfrak{Re}\left\{s_{n,m}e^{-j\phi_{n,m}}\right\} + w(t),$$

where $j = \sqrt{-1}$ denotes the imaginary number. E_s represents the average power of the emission signals; $s_{n,m}$ and $\phi_{n,m}$ are the amplitude levels and phases of the Q_l equi-probable constellation, respectively; $\mathfrak{Re}(\cdot)$ gives the real part of a complex variable. For simplicity, we consider in the analysis the received SNR can be measured [36], [37], and therefore, the transmitting power E_s will also be known.

Conditioned on the lth modulation candidate of the nth slot, the received baseband signal is further expressed as:

$$z_{n,m} = \alpha_n \sqrt{E_s} e^{-j[\phi_{n,m}(l)+\theta_n]} s_{n,m}(l) + w_{n,m}, \quad c_n = C_l. \tag{17}$$

The amplitudes levels and phases of the lth modulation constellations ($0 \le l \le |\mathcal{C}| - 1$) are given by:

$$\begin{aligned}
\text{MPSK}: \quad & \phi_{n,m}(l) = 2q\pi/Q_l, \\
& s_{n,m}(l) = 1, q = 0, 1, \cdots, Q_l - 1; \\
\text{MQAM}: \quad & \phi_{n,m}(l) = \tan^{-1}\left\{\mathfrak{Re}\left[s_{n,m}(l)\right]/\mathfrak{Im}\left[s_{n,m}(l)\right]\right\}, \\
& s_{n,m}(l) = \sqrt{\mathfrak{Re}^2[s_{n,m}(l)] + \mathfrak{Im}^2[s_{n,m}(l)]}, \\
& \mathfrak{Re}\left[s_{n,m}(l)\right], \mathfrak{Im}\left[s_{n,m}(l)\right] \in \{2q - 1 - \sqrt{Q_l}\}, \\
& q = 1, \ldots, \sqrt{Q_l},
\end{aligned}$$

where Q_l denotes the total number of equi-probable constellations of the lth candidate; $\mathfrak{Im}(\cdot)$ accounts for the image part of a complex variable. Without losing generality, five commonly used digital modulations will be considered, i.e., BPSK, QPSK, 8PSK, 16QAM and 64QAM [38].

F. Likelihood of Observations

The PDF contains useful statistic information of the observations. In classical ML schemes assuming a uniform prior, the likelihood function, conditioned on the time-varying fading gain α_n and the i.i.d. Gaussian noise, is given by:

$$\varphi_n(\mathbf{z}_n|\alpha_n) = \prod_{m=1}^{M} \sum_{C_l \in \mathcal{C}} \frac{1}{2\pi\sigma_w^2|\mathcal{C}|} \times \exp\left[-\frac{1}{\sigma_w^2}\left|z_{n,m} - \alpha_n \exp(-j\theta_n)\sqrt{E_s}e^{-j\phi_{n,m}(l)}s_{n,m}(l)\right|^2\right]. \tag{18}$$

The equivalent logarithmic-form of the likelihood function, also referred to a Log-likelihood, i.e., $\Lambda_n(\mathbf{z}_n|\alpha_n) = \log\varphi_n(\mathbf{z}_n|\alpha_n)$, is usually used in practice [18], as in (19).

$$\begin{aligned}
&\Lambda_n(\mathbf{z}_n|\alpha_n) \\
&= \frac{1}{M}\sum_{m=1}^{M} \ln \sum_{C_l \in \mathcal{C}} \exp\left[-\frac{1}{\sigma_w^2}\left|z_{n,m} - \alpha_n \exp(-j\theta_n)\right.\right. \\
&\quad \left.\left. \times \sqrt{E_s}\exp[-j\phi_{n,m}(l)]s_{n,m}(l)\right|^2\right] - \ln\left(2\pi\sigma_w^2|\mathcal{C}|\right).
\end{aligned} \tag{19}$$

It has been found that, when unknown fading parameters (e.g., α_n) are taken into accounts, the performance of the static ML schemes will be degraded seriously [6], [25]. In order to combat this realistic challenge, many schemes have been developed, e.g., ALRT [6] and MOM [25]. As far as both the implementation complexity and the recognition performance are concerned, most existing techniques may be ineffective in the context of time-varying fading channels.

In the literature, the unknown parameters will be either estimated (e.g., the MOM scheme) or marginalized (e.g., the ALRT method). The calculation of high-order moments or the marginal integration on likelihood densities, nevertheless, will be computationally extensive and may even prohibitive. Besides, such existing schemes (e.g., the MOM scheme or Hybrid LRT schemes), which requires a long observation sequence to derive accurate estimations, are *not* adaptive to the time-varying fading channels [22], [25].

More importantly, the dynamic property behind time-correlated fading channels should be exploited to further promote the performance. As demonstrated, the proper exploitation of such dynamics (or temporal memories) will result in the remarkable improvement on estimation/recognition performances [29], [33], [41]. Both the MOM methods and ALRT techniques, unfortunately, ignore the non-uniform prior and underuse the time-correlated fading channels. By focusing only on instantaneous statistical behaviors of fading (e.g., statistic PDFs or high-order moments), most existing ML approaches fail to track time-varying fading states sequentially and cannot utilize the underlying dynamics thoroughly.

G. Pilot-Assisted Channel Estimations

In some specific scenarios (e.g., a homogeneous LTE network), the *a priori* information (such as the pilot sequence) can be utilized to estimate the time-varying fading channel before the recognition process. In contrast to blind approaches, then the maximum likelihood estimator (MLE) will be suggested with the assistance of pilot [40].

Given the training sequence $\mathbf{b} = [b_1 b_2 \cdots b_{N_c}]^H$ of the length N_c, the received signal vector is then given by $\mathbf{z}_n = \mathbf{B}\boldsymbol{\alpha}_n + \mathbf{w}_n$, where the diagonal signal matrix is specified by $\mathbf{B} = \mathrm{diag}\{b_1\ b_2\ \cdots\ b_{N_c}\}$ and the channel response is $\boldsymbol{\alpha}_n = [\alpha_{n,1}, \ldots, \alpha_{n,N_c}]^T$. If define $\boldsymbol{\Sigma}_w$ as an $N_c \times N_c$ diagonal matrix with its elements are all equal to σ_w^2, it is easily found that the conditional PDF of $\mathbf{z}_n$ is Gaussian distributed, i.e., $p(\mathbf{z}_n|\boldsymbol{\alpha}_n) \rightarrow \mathcal{CN}(\mathbf{B}\boldsymbol{\alpha}_n, \boldsymbol{\Sigma}_w)$. Thus, the MLE of $\boldsymbol{\alpha}_n$ should maximize the likelihood density $p(\mathbf{z}_n|\boldsymbol{\alpha}_n)$, i.e.,

$$\begin{aligned}\hat{\boldsymbol{\alpha}}_n^{\mathrm{MLE}} &= \arg \max_{\boldsymbol{\alpha}_n \in \mathbb{C}^{N_c}} p(\mathbf{z}_n|\boldsymbol{\alpha}_n) \\ &= \arg \min_{\boldsymbol{\alpha}_n \in \mathbb{C}^{N_c}} (\mathbf{z}_n - \mathbf{B}\boldsymbol{\alpha}_n)^H \boldsymbol{\Sigma}_w^{-1} (\mathbf{z}_n - \mathbf{B}\boldsymbol{\alpha}_n). \end{aligned} \tag{20}$$

With the above linear and Gaussian model, the MLE of unknown fading channel is given by:

$$\hat{\boldsymbol{\alpha}}_n^{\mathrm{MLE}} = (\mathbf{B}^H \boldsymbol{\Sigma}_w^{-1} \mathbf{B})^{-1} \mathbf{B}^H \boldsymbol{\Sigma}_w^{-1} \mathbf{z}_n, \tag{21}$$

where $(\cdot)^H$ represents the conjugate transpose. Note that, for the considered quasi-static (or slowly-varying) fading channels, we may have $\alpha_{n,l} = \alpha_n$. Based on (21), therefore the estimation of slow-varying fading channel is finally derived from $\hat{\alpha}_n = \frac{1}{N_c} \times \sum_{l=1}^{N_c} \hat{\alpha}_{n,l}$.

III. Dynamical Modulation Recognition

In order to conform to the formulated DSM of (2)–(4), the emphasis of the following analysis is put primarily on the time-varying fading gain α_n, provided the channel phase θ_n has been known by a recognizer preliminarily. Note that, however, the channel phase θ_n will usually remain unknown, as in most realistic applications. Fortunately, the generalization of the new DSM and the designed estimation scheme to unknown channel phases, as seen from the subsequent Section III-C, is straightforward in practice.

A. Sequential Map

In contrast to classical ML-based approaches, a sequential MAP paradigm is developed for adaptive modulation recognitions, by fully exploiting the dynamic behaviors of variant channel. Our sequential Bayesian recognizer will depend on the joint posterior density

$$(\hat{\alpha}_n, \hat{c}_n) = \arg \max_{\alpha_n \in \mathcal{A}, c_n \in \mathcal{C}} p\left(\alpha_n, c_n | \mathbf{z}_{0:n}, \hat{\alpha}_{0:n-1}, \hat{c}_{0:n-1}\right), \tag{22}$$

where $c_{0:n} \triangleq \{c_0, c_1, \cdots, c_n\}$ accounts for the trajectory of modulation candidates until the nth slot; $\alpha_{0:n}$ and $\mathbf{z}_{0:n}$ represent two trajectories of time-varying fading gains and received observations, respectively.

To facilitate the derivation of the sequential MAP mechanism, two hidden states are denoted by one single state vector, i.e., $\mathbf{X}_n \triangleq \{c_n, \alpha_n\} \in \mathcal{C} \times \mathcal{A} \subset \mathbb{R}^2$. As the effective tool, the Bayesian *predict-and-update* scheme is suggested [41], [42], which recursively derives the posterior distribution of $\mathbf{X}_n$ on reception of $\mathbf{z}_n$.

1) Predict Stage: Let $p_{n-1|n-1}(\mathbf{X}_{n-1}|\mathbf{z}_{1:n-1})$ be the posterior density of the transmission slot $n-1$. Given the 1st-order Markov chain and *a priori* density $p(\mathbf{X}_0)$, the 1st-step prediction density $p_{n|n-1}(\mathbf{X}_n|\mathbf{z}_{1:n-1})$ is then obtained via the Chapman-Kolmogorov equation:

$$\begin{aligned} p_{n|n-1}(\mathbf{X}_n|\mathbf{z}_{1:n-1}) = \int_{\mathbf{X}_{n-1} \subset \mathbb{R}^2} & \phi_{n|n-1}(\mathbf{X}_n|\mathbf{X}_{n-1}) \\ & \times p_{n-1|n-1}(\mathbf{X}_{n-1}|\mathbf{z}_{1:n-1}) d\mathbf{X}_{n-1}. \end{aligned} \tag{23}$$

Here, the traditional density $\phi_{n|n-1}(\mathbf{X}_n|\mathbf{X}_{n-1})$ is related with different mapping rules between fading gains and modulation candidates. For an S2S map, it becomes:

$$\begin{aligned} &\phi_{n|n-1}(\mathbf{X}_n|\mathbf{X}_{n-1}) \\ &\quad \triangleq p_{n|n-1}\{c_n = T(\alpha_n), \alpha_n = \mathsf{A}_k | c_{n-1} = T(\alpha_{n-1}), \\ &\quad\quad \alpha_{n-1} = \mathsf{A}_{k'}\} \\ &\quad = p_{n|n-1}(\alpha_n = \mathsf{A}_k | \alpha_{n-1} = \mathsf{A}_{k'}). \end{aligned} \tag{24}$$

For another M2S map, the above transitional density is still equivalent to that of dynamic fading states, i.e., [see (25a)–(25b) at the bottom of the page].

2) Update Stage: Based on the current measurement $\mathbf{z}_n$ and the likelihood density $\varphi_n(\mathbf{z}_n|\mathbf{X}_n)$, the Bays update will be applied to refine the predicted density, i.e.,

$$p_{n|n}(\mathbf{X}_n|\mathbf{z}_{1:n}) = \frac{\varphi_n(\mathbf{z}_n|\mathbf{X}_n) p_{n|n-1}(\mathbf{X}_n|\mathbf{z}_{1:n-1})}{\int_{\mathbf{X}_n} \varphi_n(\mathbf{z}_n|\mathbf{X}_n) p_{n|n-1}(\mathbf{X}_n|\mathbf{z}_{1:n-1}) d\mathbf{X}_n}. \tag{26}$$

With the above predict-and-update propagation, the target joint density $p_{n|n}(\mathbf{X}_n|\mathbf{z}_{1:n})$ will be computed sequentially. Note that, here the likelihood $\varphi_n(\mathbf{z}_n|\mathbf{X}_n)$ actually denotes $\varphi_n(\mathbf{z}_n|\mathbf{X}_n, \hat{\theta}_n)$, i.e., assume the unknown phase θ_n has been acquired preliminarily.

It is noteworthy that, with the DSMC model, the marginal integrations in (23)–(26) can be replaced by summations. Even so, the computation complexity may still be intensive [18]. The main reason lies in that, as noted from (16), the computational complexity of calculating all kinds of likelihoods is proportional to $\mathcal{O}(LMQ_{\max})$, where $Q_{\max}$ is the maximum number of L types modulation constellations, i.e., $Q_{\max} = \max\{Q_l, 0 < l$

$$\phi_{n|n-1}(\mathbf{X}_n|\mathbf{X}_{n-1}) = \begin{cases} p_{n|n-1}(\alpha_n = \mathsf{A}_k | \alpha_{n-1} = \mathsf{A}_{k'}), & \alpha_n \in T^{-1}(c_n = C_l), \quad \text{(25a)} \\ 0, & \alpha_n \notin T^{-1}(c_n = C_l). \quad \text{(25b)} \end{cases}$$

$\leq L-1\}$. For (26), the total computation will further become $\mathcal{O}\{\max(M,K)\times LMQ_{\max}\}$.

In the following section, we will simplify the above two-step recursive estimation, with a main objective of alleviating the computational burden while insignificantly deteriorating the recognition performance.

B. Simplified Algorithms

Since different mapping rules may affect the estimation strategies of two hidden states c_n and α_n, in the following we will study the S2S and M2S mapping rules, respectively.

1) S2S Map: During a simple S2S map, as each fading state corresponds to one single modulation candidate, the predict process may be reformulated to:

$$\begin{aligned} p_{n|n-1}(\mathbf{X}_n|\mathbf{z}_{1:n-1}) &\overset{(a)}{=} p_{n|n-1}(\alpha_n|\mathbf{z}_{1:n-1}),\\ &= \sum_{\alpha_{n-1}\in\mathcal{A}} p_{n|n-1}(\alpha_n|\alpha_{n-1})p_{n-1|n-1}(\alpha_{n-1}|\mathbf{z}_{1:n-1}). \end{aligned} \tag{27}$$

Here, (a) is premised on (24). Note that, in the previous time slot $(n-1)$, the optimal estimation of fading gains has been derived by maximizing the posterior probability, i.e.,

$$\hat{\alpha}_{n-1} = \arg\max_{\alpha_{n-1}\in\mathcal{A}} p_{n-1|n-1}(\alpha_{n-1}|\mathbf{z}_{1:n-1}). \tag{28}$$

In high SNR regions, it may be further assumed that the posterior term $p_{n-1|n-1}(\hat{\alpha}_{n-1}|\mathbf{z}_{1:n-1})$ is dramatically larger than the others, i.e.,

$$\begin{aligned} &p_{n-1|n-1}(\hat{\alpha}_{n-1}|\mathbf{z}_{1:n-1})\\ &\qquad\gg p_{n-1|n-1}(\alpha_{n-1}\in\mathcal{B}_{n-1}|\mathbf{z}_{1:n-1}), \end{aligned} \tag{29}$$

where $\mathcal{B}_{n-1} \triangleq \mathcal{A}-\{\hat{\alpha}_{n-1}\}$ accounts for the difference set between $\mathcal{A}$ and $\{\hat{\alpha}_{n-1}\}$. I.e., for any element $x\in\mathcal{B}_{n-1}$, we have $x\in\mathcal{A}$ but $x\neq\hat{\alpha}_{n-1}$. Thus, we may approximate the predict term in (23) by:

$$\begin{aligned} &p_{n|n-1}(\mathbf{X}_n|\mathbf{z}_{1:n-1})\\ &\overset{(a)}{\simeq} p_{n|n-1}(\alpha_n|\hat{\alpha}_{n-1})p_{n-1|n-1}(\hat{\alpha}_{n-1}|\mathbf{z}_{1:n-1}),\\ &\overset{(b)}{\simeq} p_{n|n-1}(\alpha_n|\hat{\alpha}_{n-1}). \end{aligned} \tag{30}$$

Here, (a) holds due to (24) and (29), and (b) is based on the approximation of $p_{n-1|n-1}(\hat{\alpha}_{n-1}|\mathbf{z}_{1:n-1}) \approx 1$, by further checking (29) and the constraint relationship $\sum_{\alpha_n\in\mathcal{A}} p_{n-1|n-1}(\hat{\alpha}_{n-1}|\mathbf{z}_{1:n-1})\equiv 1$.

Similarly, the update process will be rewritten as:

$$\begin{aligned} &p_{n|n}(\mathbf{X}_n|\mathbf{z}_{1:n})\\ &= \frac{\varphi_n[\mathbf{z}_n|\alpha_n, c_n=T(\alpha_n)]\times p_{n|n-1}(\mathbf{X}_n|\mathbf{z}_{1:n-1})}{\sum_{\alpha_n\in\mathcal{A}}\varphi_n[\mathbf{z}_n|\alpha_n, c_n=T(\alpha_n)]\times p_{n|n-1}(\mathbf{X}_n|\mathbf{z}_{1:n-1})}. \end{aligned} \tag{31}$$

It is seen from (26) or (31) that the resulting term of the denominator, i.e., $\varphi_n(\mathbf{z}_n|\mathbf{z}_{1:n-1})$, will become a constant with regards to α_n, as it has marginalized α_n out, which is naturally independent of different feasible states α_n. Further combing the predict term of (30), we may simplify the updated posterior density into:

$$\begin{aligned} &p_{n|n}(\mathbf{X}_n|\mathbf{z}_{1:n})\\ &\quad\propto \varphi_n[\mathbf{z}_n|\alpha_n, c_n=T(\alpha_n)]\times p_{n|n-1}(\mathbf{X}_n|\mathbf{z}_{1:n-1}), \quad (32)\\ &\quad\simeq \varphi_n[\mathbf{z}_n|\alpha_n, c_n=T(\alpha_n)]\times p_{n|n-1}(\alpha_n|\hat{\alpha}_{n-1}). \quad (33) \end{aligned}$$

There are two points worth noting in the above simplified predict process in (30) and the update process in (33).

First, although the term $p_{n|n-1}(\alpha_n|\hat{\alpha}_{n-1})$ in (30) specifies a simple transitional density, it still can be regarded as the *sub-optimal* predicting density, as it essentially involves both the previously posterior information and *a priori* densities. The main difference is that, rather than involving the posterior density $p_{n-1|n-1}(\alpha_{n-1}|\mathbf{z}_{1:n-1})$ explicitly, the simplified term has incorporated the posterior estimation (e.g., $\hat{\alpha}_{n-1} = \arg\max_{\hat{\alpha}_{n-1}\in\mathcal{A}} p_{n-1|n-1}(\alpha_{n-1}|\mathbf{z}_{1:n-1})$) of the slot $(n-1)$ effectively into the slot n. Thus, the simplified predict and update densities are still propagated sequentially and, therefore, the recognition performance will not be remarkably compromised (especially in high SNRs).

Second, compared with the classical two-stage recursive propagation of (23) and (26), the computational complexity of the simplified process is reduced remarkably. For the assumed 1st-order DSMC (i.e., $|k_2-k_1|\leq 1$), the complexity can be measured by $\mathcal{O}(LMQ_{\max})$. That is, from (28) and (30), the likelihood functions will be computed by three times in each recursively updating procedure. In comparison, the complexity of an ALRT scheme is about $\mathcal{O}(KLMQ_{\max})$, in order to marginalize the unknown fading gain.

Based on the log-likelihood, the updated posterior density will be re-formatted to:

$$\begin{aligned} &p_{n|n}(\mathbf{X}_n|\mathbf{z}_{1:n})\\ &\quad\propto \ln\left\{\varphi_n[\mathbf{z}_n|\alpha_n, c_n=T(\alpha_n)]\times p_{n|n-1}(\alpha_n|\mathbf{z}_{1:n-1})\right\},\\ &\quad\simeq \Lambda(\mathbf{z}_n|\alpha_n)+\ln p_{n|n-1}(\alpha_n|\hat{\alpha}_{n-1}). \end{aligned}$$

With the sequentially propagated posterior density, the MAP estimations of the variant fading state and the associated modulation candidate are obtained via (34)-(35), shown at the bottom of the page.

2) M2S Map: The estimation scheme developed for the S2S mapping, in practice, can be referred to a *fading-driven* method, which can be conveniently applied to an M2S map. It is seen that such a fading-driven method will firstly estimate

$$\boxed{\begin{aligned} &\hat{\alpha}_n = \arg\max_{\alpha_n\in\mathcal{A}}\left\{\Lambda(\mathbf{z}_n|\alpha_n)+\ln p_{n|n-1}(\alpha_n|\hat{\alpha}_{n-1})\right\},\\ &\hat{c}_n = T(\hat{\alpha}_n). \end{aligned}} \quad \begin{aligned}(34)\\(35)\end{aligned}$$

the fading channel α_n, and then directly obtain the modulation scheme from $c_n = T(\alpha_n = \mathsf{A}_k)$. Since multiple fading states correspond to one modulation candidate in the other M2S rule, a different estimation strategy may be suggested intuitively. This 2nd simplified scheme, referred to the *goal-oriented* method, will instead estimate the unknown modulation candidate directly.

Based on the prior transitional density $p_{n|n-1}(c_n|c_{n-1})$ and the previous information, firstly the predict process can be derived via:

$$p_{n|n-1}(c_n|\mathbf{z}_{1:n-1}) = \sum_{c_{n-1}\in\mathcal{C}} p_{n|n-1}(c_n|c_{n-1}) \times p_{n-1|n-1}(c_{n-1}|\mathbf{z}_{1:n-1}). \quad (36)$$

With the estimated candidate $\hat{c}_{n-1}$ and the similar simplifications in high SNRs, i.e., (27)–(30), the predicted density is calculated by:

$$\begin{aligned} p_{n|n-1}(c_n|\mathbf{z}_{1:n-1}) &\simeq p_{n|n-1}(c_n|\hat{c}_{n-1}) \times p_{n-1|n-1}(\hat{c}_{n-1}|\mathbf{z}_{1:n-1}), \quad (37)\\ &\simeq p_{n|n-1}(c_n|\hat{c}_{n-1}). \quad (38)\end{aligned}$$

Thus, the predicting process will be realized effectively by (16). Before elaborating on the updating strategy, we may firstly consider a practical scenario that is of significance to the designing of the goal-oriented method.

Assume $\alpha_{n-1} \in T^{-1}(C_l) = \{\mathsf{A}_k, \mathsf{A}_{k-1}\}$ and $\alpha_n \in T^{-1}(C_{l+1}) = \{\mathsf{A}_{k+1}, \mathsf{A}_{k+2}\}$ (i.e., $|\mathcal{A}_l| = |\mathcal{A}_{l+1}| = 2$) and the fading state of the $(n-1)$th time slot has been estimated, e.g., $\hat{\alpha}_{n-1} = \mathsf{A}_k$ and $\hat{c}_{n-1} = C_l$. In the nth time slot, then the prior of residing in $c_n = C_l$ is calculated via $p(c_n = C_l|\hat{c}_{n-1} = C_l) = p(\alpha_n = \mathsf{A}_k|\hat{\alpha}_{n-1} = \mathsf{A}_k) + p(\alpha_n = \mathsf{A}_{k-1}|\alpha_{n-1} = \mathsf{A}_k)$, while the prior transiting to C_{l+1} is only given by $p(c_n = C_{l+1}|\hat{c}_{n-1} = C_l) = p(\alpha_n = \mathsf{A}_{k+1}|\hat{\alpha}_{n-1} = \mathsf{A}_k)$. Given the assumed M2S map and the 1st-order slowly varying DSMC, the *a priori* of the first transition (i.e., C_l) will probably dominate the second one [32]. In the real situation $c_n = C_{l+1}$, then the error recognition $\hat{c}_n = C_l$ will occur easily, especially when the SNR is low and the likelihoods are not sufficiently discriminate to prevent the wrong prior transition of $\hat{c}_n = C_l$. What's more, the above wrong recognition, accompanying its reacting prior transitions, will be propagated in subsequent slots, resulting in serious performance degradations. It is found that, as a consequence, the original update equation seems to be not appropriate anymore. Note that, this realistic difficulty is aroused mainly by the coupling relation of the M2S map.

In order to overcome the problem and restrict erroneous transitions during the estimation process, we further suggest a modified likelihood to replace classical likelihoods, which is defined as:

$$\tilde{\varphi}(\mathbf{z}_n|c_n, \hat{c}_{n-1}, \hat{\alpha}_{n-1}) \triangleq \frac{\sum_{\alpha_n\in\mathcal{A}} \varphi_n(\mathbf{z}_n|c_n, \alpha_n)p(\alpha_n|\hat{\alpha}_{n-1})}{\sum_{\alpha_n\in\mathcal{A}} p(\alpha_n|\hat{\alpha}_{n-1})}. \quad (39)$$

It is seen from the (39) that this weighted likelihood $\tilde{\varphi}(\mathbf{z}_n|c_n, \hat{c}_{n-1}, \hat{\alpha}_{n-1})$ may be viewed essentially as an *averaged* posterior density, as it involves both the prior transitional probabilities $p(\alpha_n|\hat{\alpha}_{n-1})$ and the likelihoods $\varphi_n(\mathbf{z}_n|c_n, \alpha_n)$.

By resorting to $\tilde{\varphi}_n(\mathbf{z}_n|c_n, \hat{c}_{n-1}, \hat{\alpha}_{n-1})$, the updating process may be implemented alternatively. That is, the estimation of unknown modulation candidate, relying on the *averaged* posterior density, will be directly derived via:

$$\boxed{\hat{c}_n = \arg\max_{c_n\in\mathcal{C}} \tilde{\varphi}_n(\mathbf{z}_n|c_n, \hat{c}_{n-1}, \hat{\alpha}_{n-1}),} \quad (40)$$

Subsequently, the fading state α_n will be recovered by maximizing the posterior density that is conditioned on the above estimated candidate, i.e.,

$$\begin{aligned} \hat{\alpha}_n &= \arg\max_{\alpha_n\in\mathcal{A}} p_{n|n}(\alpha_n|\mathbf{z}_{1:n}, \hat{c}_{1:n}), \quad (41)\\ &= \arg\max_{\alpha_n\in\mathcal{A}} \varphi_n(\mathbf{z}_n|\alpha_n, \hat{c}_n)p_{n|n}(\hat{c}_n|\alpha_n)p_{n|n-1}(\alpha_n|\hat{c}_{n-1}), \quad (42)\\ &\propto \arg\max_{\alpha_n\in T^{-1}(\hat{c}_n)} \left\{\Lambda(\mathbf{z}_n|\alpha_n, \hat{c}_n) + \ln p_{n|n-1}(\alpha_n|\hat{\alpha}_{n-1})\right\}. \quad (43)\end{aligned}$$

Compared with the 1st fading-driven method, i.e., first acquire the fading gain and then estimate the modulation candidate, the 2nd goal-oriented scheme may be less completive in high SNR regions, based on the following two considerations. First, the map from fading states to modulation candidates is always single-to-single (even for M2S), the direct estimation of fading states, therefore, can be viewed as a *joint* Bayesian framework. The 2nd scheme, in contrast, estimates modulation candidates and fading states separately. Second, in the goal-oriented scheme, both the estimation of modulation candidates and fading states will depend on the marginalization of the other one. This procedure may lose some information, when compared with the joint estimation scheme. Note that, nevertheless, the 2nd scheme may be more competitive in low SNRs region, as the weighted likelihood of (39) may prevent the erroneous transitions and the estimation error propagations.

C. Channel Phase Estimation

As an intuitive approach, by assuming the unknown phase θ_n is distributed *uniformly*, a marginalization process may be suggested to mitigate its deteriorating effects, as in the ALRT scheme [22], [39]. Or alternatively, another dynamic equation, i.e., $\theta_n = \Theta(\theta_{n-1})$, can be introduced to characterize the unknown channel phase, which would be then estimated by using a similar Bayesian framework developed for time-varying fading gains.

As far as the additional complexity is concerned, the ALRT approach seems to be infeasible when it is applied to estimate the *continuous* channel phase θ_n. The introduction of another dynamic function, on the other hand, will also be of little effects, as the channel phase is usually not time-correlated, i.e., $\theta_n|\theta_{n-1} \sim \mathrm{U}(0, 2\pi]$. Thus, in this work an MOM-based pre-estimation scheme is suggested to acquire unknown channel phases, as in [22]. Given the observation sequence $\{z_{n,m}\}$, then the channel phases will be estimated via:

$$\hat{\theta}_n^{\mathrm{M-PSK}} = M_c^{-1} \times \mathtt{ang}\left(\sum_{m=1}^{M} z_{n,m}^{M_c}\right), \quad (44)$$

$$\hat{\theta}_n^{\mathrm{M-QAM}} = 4^{-1} \times \mathtt{ang}\left(-\sum_{m=1}^{M} z_{n,m}^{4}\right). \quad (45)$$

Here, $\mathtt{ang}(x)$ accounts for the phase of a complex variable x. It is noted that the integer $M_c = 2^{m_0}$ is typically larger than 4 (i.e., $m_0 \geq 2$) [22]. For example, it is configured to 8

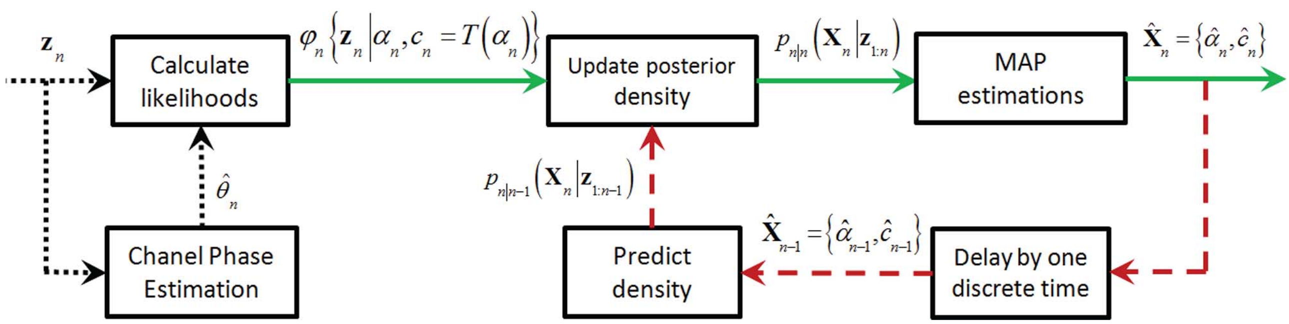


Fig. 2. The schematic structure of the proposed Bayesian approach for dynamic modulation recognitions. Note that, the *green* solid-lines denote the update process, while the *red* dash-lines correspond to the predict process. The dotted lines amount for the pre-estimation of unknown channel phases.

for the 8-PSK signal. In practice, the modulation scheme will usually remain unknown to a recognizer. Thus, it has to derive multiple estimations of channel phases, i.e., $\hat{\theta}^{\mathrm{M-PSK}}_{M_c}$ and $\hat{\theta}^{\mathrm{M-QAM}}_{M_c}$, by assuming various feasible MPSK or MQAM constellations. Then, the likelihood of (19) will be evaluated for multiple times, given different estimation candidates of phases and fading gains. Finally, the complex fading channel accompanying the adopted modulation index will be acquired via the designed MAP scheme.

Noted that, further considering the pre-estimation of channel phase, the complexity of the proposed scheme will become $\mathcal{O}\{K_1 \times [LMQ_{\max} + M\log_2(M_c)]\}$. Here, K_1 denotes the number of multiple channel phases and, we have $K_1 \leq K$ in our analysis.

D. Implementations

Based on the above elaborations, the simplified dynamic modulation recognition algorithm will be implemented based on the following four steps. Here, we only summarize the 1st fading-driven method, while the 2nd scheme may be implemented similarly.

Step 1: The unknown phase θ_n of fading channels is firstly derived from (44)–(45), relying on an MOM method.

Step 2: With the MAP estimations of the time $n-1$ (e.g., $\hat{\alpha}_{n-1}$ and $\hat{c}_{n-1}$), the predict term is obtained via (30).

Step 3: Conditioned on the estimated channel phase $\hat{\theta}_n$, the related likelihood densities are evaluated, given the current observation $\mathbf{z}_n$. With the help of the predict density, then the update density $p_{n|n}(\alpha_n|\alpha_{0:n-1}, \mathbf{z}_{0:n-1})$ will be derived based on (32).

Step 4: Finally, the other two hidden states of the time n will be estimated by maximizing the updated posterior density based on (34) and (35). The estimated states will further be reserved for the next time slot.

Taking the fading-driven method still for example, a schematic structure is demonstrated by Fig. 2.

IV. Numerical Simulations

In this section, both the recognition and estimation performance of the proposed Bayesian approach will be investigated, in the context of realistic time-varying fading channels. Without losing generality, in the simulation the Rayleigh fading with a distribution variance $\sigma^2 = 0.1$ is assumed. For the FSMC model, the number of representative states is K=6 and, therefore, the stationary residue probability is $\pi_k = 1/6$. Given a typical configuration of the Doppler frequency shift f_D or the statical length L_a (e.g., $L_a = 10, 20$ or 50), then K fading states A_k accompanying the TPM $\mathbf{P}_{M\times M}$ will be calculated from (8)-–(11). Two map rules are considered, i.e., S2S and M2S. In an M2S map, two successive fading regions are mapped to one single modulation candidate. As illustrated by Table I, three high-order modulation schemes are concerned, i.e., QPSK, 16QAM and 64QAM. In this situation, we have $|\mathcal{C}| = 3$ and $|\mathcal{A}_l| = 2$ for $l = 1, 2, 3$. For the other S2S map, the time-varying fading gain is split into $K = 6$ sub-regions and total $|\mathcal{C}| = 6$ modulation candidates are adopted, i.e., BPSK, QPSK, 8PSK, 16QAM, 32QAM and 64QAM.

TABLE I
Mapping Rules of Adaptive Modulation

Fading States	A_1	A_2	A_3	A_4	A_5	A_6
S2S map	BPSK	QPSK	8PSK	16QAM	32QAM	64QAM
M2S map	QPSK		16QAM		64QAM	

Rather than the recognition performance of each single candidate, an averaging recognition ratio is used, which takes L kinds of recognition ratios $\Pr(\hat{c}_n = C_l|c_n = C_l)$ into accounts, i.e.,

$$\begin{aligned} P_D &\triangleq \sum_{l=1}^{L} p(c_n = C_l) \times \Pr(\hat{c}_n = C_l|c_n = C_l) \\ &= \sum_{l=1}^{L} p(\alpha_n \in \mathcal{A}_l) \times \Pr(\hat{c}_n = C_l|c_n = C_l), \\ &= \sum_{l=1}^{L} |\mathcal{A}_l|\pi_k \times \Pr(\hat{c}_n = C_l|c_n = C_l). \end{aligned} \tag{46}$$

The other metric of our new joint MAP scheme is the mean square error (MSE) of the estimated fading gains, which is defined as:

$$\mathrm{MSE} \triangleq \frac{1}{N} \times \mathbb{E}\left\{\sum_{n=1}^{N} \frac{|\hat{\alpha}_n e^{-j\hat{\theta}_n} - \alpha_n e^{-j\theta_n}|^2}{|\alpha_n|^2}\right\}. \tag{47}$$

A. Fading-Driven or Goal-Oriented?

In the first experiment, we will study the two proposed schemes, i.e., the 1st fading-driven scheme and the 2nd goal-oriented method. In the simulation, the M2S map is concerned and

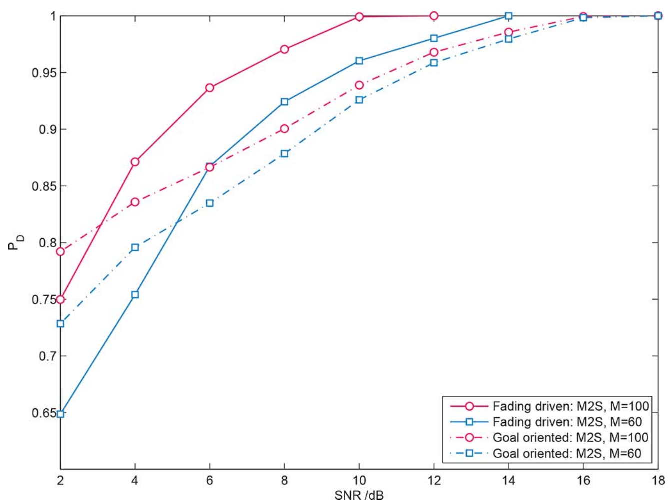


Fig. 3. The recognition performance of both the fading-driven and goal-oriented schemes. In the experiment, the M2S mapping criterion is considered and the static length is configured to $L_a = 20$.

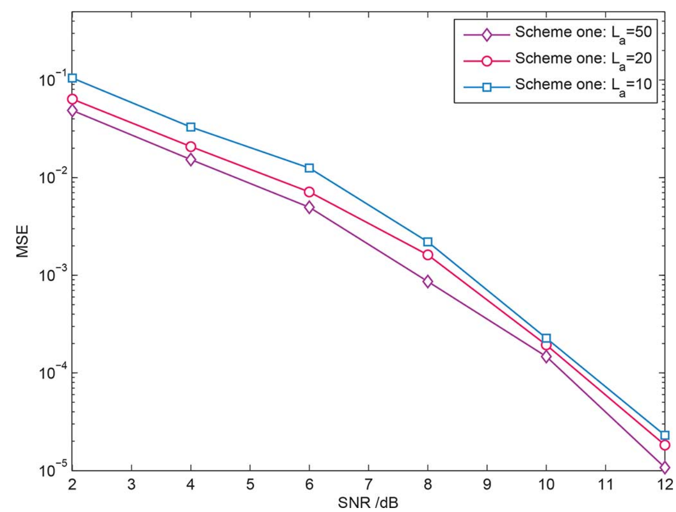


Fig. 4. The MSE of estimated fading channels under different L_a. In the experiment, the 1st fading-driven scheme is adopted and the M2S mapping rule is considered. The samples size is configured to $M = 100$.

the static length of time-varying fading is set to $L_a = 20$. Each curve is derived from 20 independent realizations of random fading channels, in which total 10000 modulated symbols are transmitted. We see from the experimental results shown by Fig. 3 that, in high SNRs (e.g., >4 dB for $M = 60$), the scheme one (i.e., fading-driven) is indeed superior to the scheme two (i.e., goal-oriented). When the averaging recognition ratio P_D is larger than 0.95, a detection gain of 2 dB can be achieved by the scheme one if the sample size is $M = 60$. As analyzed, the scheme one realizes essentially the joint estimation of unknown modulation candidates and dynamic fading states. The scheme two, despite goal-oriented, may loss some information due to the marginalization operation, which acquires less attractive performance. In low SNR regions (e.g., <4 dB), however, the scheme two seems to be more competitive. With the new weighted likelihood in (39), the error propagations aroused by the rashly advance transition may be restricted, resulting in a promoted recognition performance. We conclude that these two schemes may be combined in adaptive modulation recognitions given different realistic SNRs.

B. Different Fading Channels

In the second experiment, the effect of different fading channels is studied. In the simulation, we configure the sample size M to 100. For most slowly-varying fading channels, the Doppler frequency f_D is relatively small. Thus, three typical configurations are considered, i.e., the static length L_a is set to 10, 50 and 20, respectively. From the MSE of estimated fading channels in Fig. 4, it is seen that the estimation accuracy will be improved slightly with the increasing of a static length L_a. For example, if the SNR is 8 dB, the MSE value in the case of $L_a = 50$ is about 8.5×10^{-4}. In comparison, the MSE of $L_a = 10$ will be decreased to 2.3×10^{-3}.

In Fig. 5, the averaging recognition ratio P_D of the proposed scheme one is demonstrated. For comparative analysis, the other existing counterpart, i.e., the ALRT scheme, is also studied, as it relies similarly on likelihoods and takes also the random

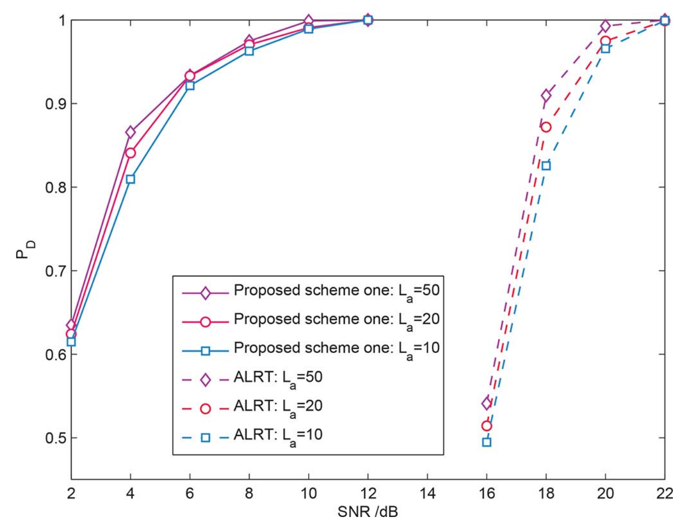


Fig. 5. The average recognition accuracy under different L_a. In the experiment, the 1st fading-driven scheme is adopted and the M2S mapping rule is considered. The samples size is configured to $M = 100$.

fading channels into account. During the ALRT scheme, by directly marginalizing out the unknown fading gain according to its prior PDF, the time-correlated property of fading channels (or adjacent modulation candidates) is ignored. Thus, it can hardly exploit the underlying dynamics of fading channels. Like most ML-based recognition approaches that are extremely sensitive to variant fading effects, the recognition performance of the ALRT method will be deteriorated. Premised on the formulated DSM, in contrast, the proposed joint estimate scheme will track fading channels adaptively. By fully utilizing the dynamics of time-varying fading gains, the recognition performance will be improved. From the numerical results, it is noted that the proposed scheme (e.g., the scheme one), relying on the Bayesian estimation paradigm, can acquire 8–9 dB gain in realistic time-varying fading channels. Besides, the recognition performance will not be affected by different fading rates (or the static length L_a), since the varying fading states have been accurately estimated.

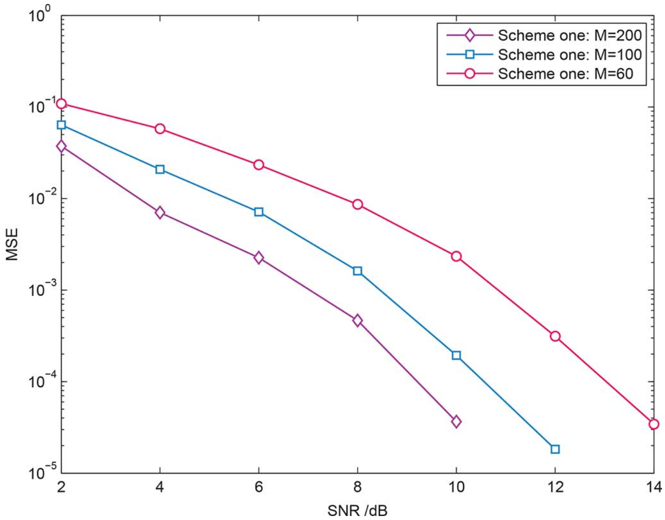


Fig. 6. The MSE of estimated fading channels under different M. The 1st fading-driven scheme is adopted and the M2S mapping rule is considered. The static length of fading channel is configured to $L_a = 20$.

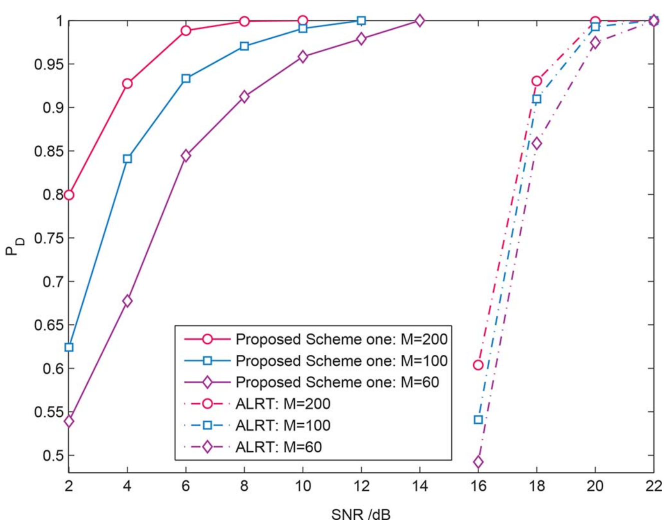


Fig. 7. The average recognition accuracy under different M. The 1st fading-driven scheme is adopted and the M2S mapping rule is considered. The static length of fading channel is configured to $L_a = 20$.

C. Different Sampling Sizes

In this experimental simulation, we will evaluate the performance of the joint recognition and estimation scheme under different sample sizes M. The static length is configured to $L_a = 20$. It is easy to understand that, with more observation samples, more information can be utilized to improve the accuracy of channel tracking and modulation recognition. As shown by Figs. 6 and 7, for the designed scheme, the performance of both MSE and recognition accuracy will be improved with an increasing of the sample size M. Due to the corruptions of variant fading gains, however, the recognition performance of the existing ALRT methods will not be improved by simply increasing the sample size M. Note that, in practice, a compromise should also be made between the performance and the computational complexity. Except for the prolonged sampling time, a larger sample size M may also increase the computational complexity of likelihood significantly (recall that it is practically proportional to M).

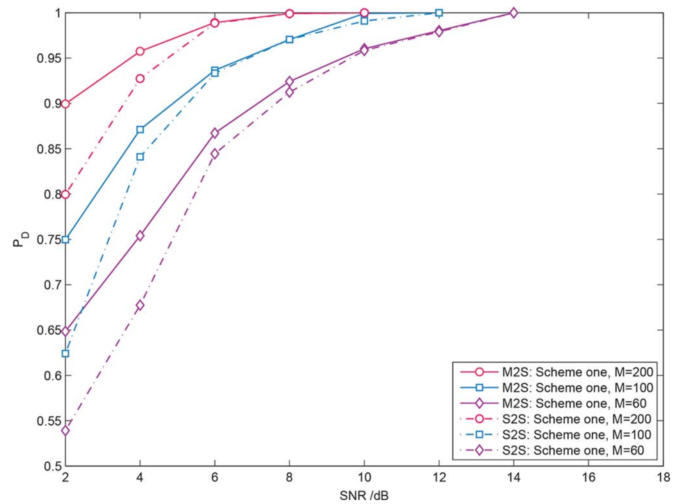


Fig. 8. The average recognition accuracy in both the S2S and M2S mapping rules. The 1st fading-driven scheme is adopted and the sample size is configured to $M = 100$.

D. Different Map Schemes

In the experiment, the recognition performances under both S2S and M2S mapping rules are investigated. The numerically derived recognition performances have been plotted in Fig. 8. Here, the static length is configured to $L_a = 20$ and the sample size is configured to $M = 100$.

It is seen that, compared with an S2S map of $L = 6$, the average recognition accuracy of the M2S mapping rule $(L = 3)$ will be promoted. This is mainly because, with an S2S map, the coupling relationship between the unknown fading states and modulation candidates will be closer together. Thus, the residue error on the estimated fading gains will dramatically affect the modulation recognition. In comparison, when multiple fading regions are merged into only one modulation scheme as in the M2S map, the sensitivity of recognition procedure to $|\mathcal{A}_l|$ adjacent fading states will be alleviated to some extent. From the simulation result, in low SNRs (e.g., $P_D = 0.8$ and $M = 100$), the M2S map may even achieve 1.1 dB gain compared with the other S2S map. In high SNRs (e.g., $P_D > 0.95$), the two different mapping rules, however, will acquire comparable performances, as the unknown fading gains have already been estimated accurately.

E. Different Modulation Candidates

In this simulation, the recognition accuracy of each single modulation candidate is investigated. The M2S mapping rule is considered. The static length is $L_a = 20$ and the sample size is $M = 100$. From the experimental results, in the context of adaptive modulation recognitions, the accuracy of different modulated signals will be quite different, as shown by Fig. 9. In practice, the recognition accuracy of each candidate will depend on both the *instant* SNRs and the likelihoods. The

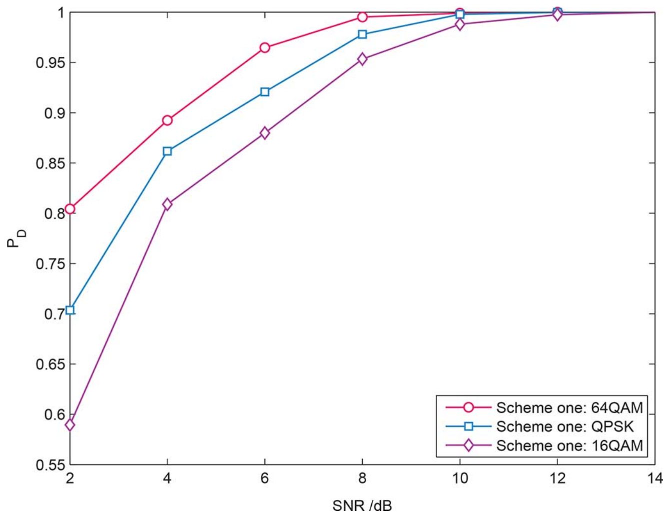


Fig. 9. The recognition accuracy of single modulation candidate. The M2S mapping rule is considered and the 1st fading-driven scheme is adopted. The sample size is configured to $M = 100$.

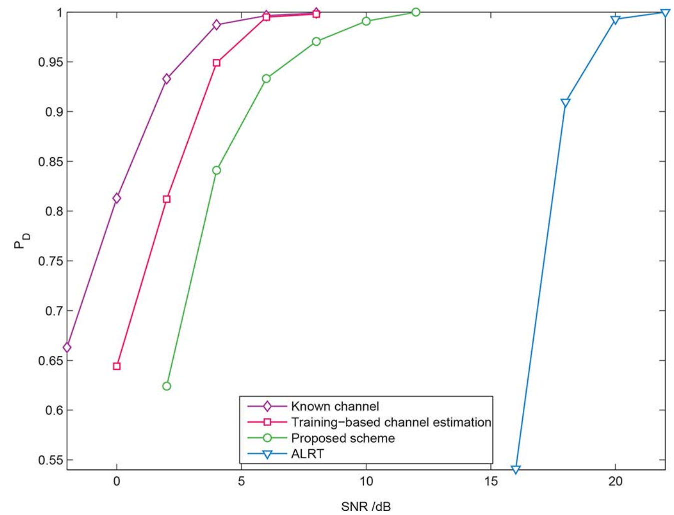


Fig. 10. The recognition performance with the training-based channel estimation and the proposed recursive MAP estimation scheme. The M2S mapping rule is considered and the 1st fading-driven scheme is adopted. The sample size is $M = 100$ and the statical length is $L_a = 20$. For comparative analysis, a benchmark curve have also been provided, by assuming the channel responses are perfectly known.

first factor is referred to the instantaneous SNR of certain time slots with specific fading gains (i.e., α_n^2). Note that, the instant SNRs are different from the averaging SNR adopted before (i.e., related with $\mathbb{E}_{\mathcal{A}}\{\alpha_n^2\}$). The second factor, i.e., the likelihood distributions, actually gives the *discrimination* with regards to other modulation schemes. Usually, different modulation candidates may have different discriminations. The recognition accuracy of single modulation mode, therefore, will be related with the fading condition, the map scheme and the specific modulation set. For the specific configuration, 64QAM may obtain the superior recognition performance.

F. Blind vs. Training-Based

In the last experiment, the recognition performance with the training-based channel estimation is investigated for the comparative analysis. A pilot (or preamble) sequence of the length $N_c = 3$ is used, i.e., $\mathbf{b} = \frac{1}{\sqrt{2}} \times [1+j \;\; 1-j \;\; -1+j]^T$. The QPSK modulation format of this training sequence is also assumed to be known by the recognizer. Thus, the time-varying fading channel will be estimated by the receiver via the data-aided ML criterion, as in (21).

The performances of three different recognition methods are plotted together in Fig. 10. It is seen that the pilot-assisted scheme will acquire the superior recognition performance, by acquiring the precise estimation of fading channels. Compared with the proposed joint estimation scheme, a detection gain of 2–3 dB would be achieved by the pilot-assisted scheme. Meanwhile, the benchmark performance with the perfectly known channel is also provided. The SNR gap between the proposed scheme and the idea case with the perfect channel knowledge is about 4–5 dB. It is noteworthy that, like the ALRT scheme, however the proposed sequential MAP scheme belongs essentially to a blind (or semi-blind) approach, which requires only *a priori* information on the fading distribution. By excluding the training sequence that may be unavailable in some realistic scenarios, the proposed new scheme, which, therefore, possesses the higher energy and frequency efficiency, may provide a great convenience to future mobile networks or other uncoordinated applications (e.g., D2D).

V. Conclusion

We investigated the adaptive modulation recognition for the emerging 5G mobile applications with time-varying fading channels. In order to characterize the mutually coupling relation between two hidden states, i.e., variant fading channels and unknown modulation scheme, a new DSM model is established, by taking the dynamic behavior of time-dependent fading channels into account. On this basis, a promising recognition scheme is designed. A major innovation of our new scheme is that, rather than marginalizing random fading effects out, it will sequentially estimate the time-varying fading gains, at the same time of recognizing the unknown adaptive modulation mode. Premised on a Bayesian inference framework, two simplified estimation algorithms are developed. Simulations validate the proposed schemes. By profoundly exploiting the underlying dynamics of time-correlated fading channels and the coupling relationship among two hidden terms, the recognition performance will be improved significantly. The formulated general DSM and the joint estimation paradigm, which may be generalized to other modulation scenarios (i.e., unknown noise variance), will enhance the link adaptation performance and may be of great promise to future mobile 5G communications.

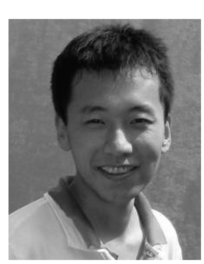
References

[1] J. G. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, and A. C. K. Soong *et al.*, “What will 5G be?,” *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

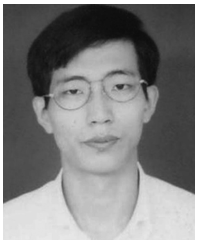
[2] J. Meng and E. H. Yang, “Constellation and rate selection in adaptive modulation and coding based on finite block-length analysis and its application to LTE,” *IEEE Trans. Wireless Commun.*, vol. 13, no. 10, pp. 5496–5508, 2014.

[3] A. Panajotovic, F. Riera Palou, and G. Femenias, “A novel adaptive UCD-based MU-MIMO scheme for IEEE 802.11ac,” in *Proc. 20th Eur. Wireless Conf. (European Wireless’14)*, Barcelona, Spain, May 14–16, 2014, pp. 1–6.

[4] A. Goldsmith and S. G. Chua, "Adaptive coded modulation for fading channels," *IEEE Trans. Commun.*, vol. 46, pp. 595–602, May 1998.
[5] A. Svensson, "An introduction to adaptive QAM modulation schemes for known and predicted channels," *Proc. IEEE*, vol. 95, no. 12, pp. 2322–2336, 2007.
[6] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: Classical approaches and new trends," *IET Commun.*, vol. 1, no. 2, pp. 137–156, Apr. 2007.
[7] K. Kim and A. Polydoros, "Digital modulation classification: The BPSK versus QPSK case," in *Proc. IEEE Military Commun. Conf. (MLCOM)*, San Diego, CA, USA, Oct. 23–26, 1988, vol. 2, pp. 431–436.
[8] B. F. Beidas and C. L. Weber, "Higher-order correlation-based classification of asynchronous MFSK signal," in *Proc. IEEE Military Commun. Conf. (MLCOM)*, McLean, VA, USA, Oct. 24th, 1996, pp. 1003–1009.
[9] P. Panagiotou, A. Anastasoupoulos, and A. Polydoros, "Likelihood ratio tests for modulation classification," in *Proc. IEEE Military Commun. Conf. (MLCOM)*, Los Angeles, CA, USA, Oct. 22th–25th, 2000, pp. 670–674.
[10] J. L. Xu, W. Su, and M. C. Zhou, "Software-defined radio equipped with rapid modulation recognition," *IEEE Trans. Veh. Technol.*, vol. 59, no. 4, pp. 1659–1667, 2010.
[11] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Commun. Surveys Tuts.*, pp. 1–18, 2014, doi: 10.1109/COMST. 2014.2319555.
[12] D. Boutte and B. Santhanam, "A hybrid ICA-SVM approach to continuous phase modulation recognition," *IEEE Signal Process. Lett.*, vol. 16, no. 5, pp. 402–405, 2009.
[13] A. K. Nandi and E. E. Azzouz, "Modulation recognition using artificial neural networks," *Signal Process.*, vol. 56, no. 2, pp. 165–175, 1997.
[14] M. Marey and O. A. Dobre, "Blind modulation classification algorithm for single and multiple-antenna systems over frequency-selective channels," *IEEE Signal Process. Lett.*, vol. 21, no. 9, pp. 1098–1102, Sep. 2014.
[15] A. Swami and B. M. Sadler, "Hierarchical digital modulation classification using cumulants," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 416–429, Mar. 2000.
[16] K. Hassan, I. Dayoub, W. Hamouda, and M. Berbineau, "Automatic modulation recognition using wavelet transform and neural networks in wireless systems," in *EURASIP J. Adv. Signal Process.*, p. 13, Article ID 532898.
[17] J. Lundén and V. Koivunen, "Automatic radar waveform recognition," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 1, pp. 124–136, 2007.
[18] D. Bai, J. Lee, S. Kim, and I. Kang, "Near ML modulation classification," in *Proc. IEEE Veh. Technol. Conf. (VTC Fall)*, Québec City, Canada, Sep. 3–6, 2012, pp. 1–5.
[19] J. A. Sills, "Maximum-likelihood modulation classification for PSK/QAM," in *Proc. IEEE Military Commun. Conf. (MLCOM)*, Atlantic City, NJ, USA, 1999, pp. 57–61.
[20] W. Wei and J. M. Mendel, "Maximum-likelihood classification for digital amplitude-phase modulations," *IEEE Trans. Commun.*, vol. 48, no. 2, pp. 189–193, 2000.
[21] A. E. El-Mahdy and N. M. Namazi, "Classification of multiple M-ary frequency-shift keying signals over a Rayleigh fading channel," *IEEE Trans. Commun.*, vol. 50, no. 6, pp. 967–974, 2002.
[22] O. A. Dobre and F. Hameed, "Likelihood-based algorithms for linear digital modulation classification in fading channel," in *Proc. Canad. Conf. Electr. Comput. Eng. (CCECE'06)*, Ottawa, ON, Canada, 2006, pp. 1347–1350.
[23] Y. C. Lin and C. J. Kuo, "Classification of quadrature amplitude modulated (QAM) signals via sequential probability ratio test (SPRT)," *Signal Process.*, vol. 60, no. 3, pp. 263–280, 1997.
[24] F. Hameed, O. A. Dobre, and D. C. Popescu, "On the likelihood-based approach to modulation classification," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 5884–5892, Dec. 2009.
[25] 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, 2011.
[26] S. Barbarossa, S. Sardellitti, and P. Di Lorenzo, "Communicating while computing: Distributed mobile cloud computing over 5G heterogeneous networks," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 45–55, 2014.
[27] B. Sklar, "Rayleigh fading channels in mobile digital communication systems I: Characterization," *IEEE Commun. Mag.*, vol. 35, no. 7, pp. 90–100, Aug. 2002.
[28] L. R. Rabiner and B. H. Juang, "An introduction to hidden Markov models," *IEEE Acoust., Speech, Signal Process. Mag.*, pp. 4–16, Jan. 1986.
[29] P. Sadeghi, R. Kennedy, P. Rapajic, and R. Shams, "Finite-state Markov modeling of fading channels: A survey of principles and applications," *IEEE Signal Process. Mag.*, vol. 25, no. 5, pp. 57–80, 2008.
[30] F. Babich and G. Lombardi, "A Markov model for the mobile propagation channel," *IEEE Trans. Veh. Technol.*, vol. 49, no. 1, pp. 63–73, Jan. 2000.
[31] B. Li, Z. Zhou, and A. Nallanathan, "Joint estimation based spectrum sensing for cognitive radios in time variant flat fading channel," in *Proc. IEEE Global Commun. Conf. (Globecom)*, Atlanta, GA, USA, Dec. 2013, pp. 1–6.
[32] H. S. Wang and N. Moayeri, "Finite-state Markov channel: A useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol. 44, no. 1, pp. 163–171, Feb. 1995.
[33] H. S. Wang and P. Chang, "On verifying the first order markovian assumption for a Rayleigh fading channel model," *IEEE Trans. Veh. Technol.*, vol. 45, no. 2, pp. 353–357, May 1996.
[34] B. Li, C. L. Zhao, M. W. Sun, and A. Nallanathan, "Spectrum sensing for cognitive radios in time-variant flat fading channels: A joint estimation approach," *IEEE Trans. Commun.*, vol. 62, no. 8, pp. 665–2680, Aug. 2014.
[35] B. Canpolat and Y. Tanik, "Performance analysis of adaptive loading OFDM under Rayleigh fading," *IEEE Trans. Veh. Technol.*, vol. 53, no. 4, pp. 1105–1115, July 2004.
[36] D. R. Pauluzzi and N. C. Beaulieu, "A comparison of SNR estimation techniques for the AWGN channel," *IEEE Trans. Commun.*, vol. 48, no. 10, pp. 1681–1691, Oct. 2000.
[37] A. Wiesel, J. Goldberg, and H. Messer-Yaron, "SNR estimation in time-varying fading channels," *IEEE Trans. Commun.*, vol. 54, no. 3, pp. 841–848, May 2006.
[38] J. G. Proakis, *Digital Communications*, 5th ed. New York, NY, USA: McGraw-Hill, 2008.
[39] O. A. Dobre, J. Zarzoso, Y. Bar-Ness, and W. Su, "On the classification of linearly modulated signals in fading channel," in *Proc. 38th Annu. Conf. Inf. Sci. Syst. (CISS 2004)*, Princeton, NJ, USA, Mar. 2004, pp. 1347–1350.
[40] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1998, Volume I: Estimation Theory.
[41] B. Li, M. W. Sun, C. L. Zhao, and A. Nallanathan, "Energy detection based spectrum sensing for cognitive radios over time-frequency doubly selective fading channels," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 402–417, Jan. 2015.
[42] O. Cappó, E. Moulines, and T. Rydén, *Inference in Hidden Markov Models*. New York, NY, USA: Springer, 2005.
[43] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. New York, NY, USA: Academic, 1970.



Bin Li received the Bachelor's degree in electrical information engineering from Beijing University of Chemical Technology (BUCT) in 2007, the Ph.D. degree in communication and information engineering from Beijing University of Posts and Telecommunications (BUPT) in 2013. He joined BUPT since 2013, as a lecture of the School of Information and Communication Engineering (SCIE). His current research interests are focused on statistical signal processing algorithms for wireless communications, e.g., ultra-wideband (UWB), wireless sensor networks, millimeter-wave (mm-Wave) communications and cognitive radios (CRs). He has published more than 60 journal and conference papers. He received 2011 ChinaCom Best Paper Award, 2010 and 2011 BUPT Excellent Ph.D. Student Award Foundation. He served as the regular reviewer for IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON SIGNAL PROCESSING and IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.

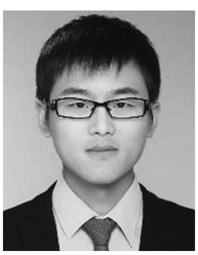


Shenghong Li received the B.S. and the M.S. degrees in electrical engineering from Jilin University of Technology, China, in 1993 and 1996, respectively, and received the Ph.D. degree in radio engineering from Beijing University of Posts and Telecommunications (BUPT), China, in 1999. Since 1999, he has been working in Shanghai Jiao Tong University (SJTU), China, as Research Fellow, Associate Professor and Professor, successively. In 2010, he worked as Visiting Scholar in Nanyang Technological University, Singapore. His research interests include information security, signal and information processing, cognitive radio networks, artificial intelligence.

He published more than 80 papers, co-authored four books, and holds ten granted patents. In 2003, he received the 1st Prize of Shanghai Science and Technology Progress in China. In 2006 and 2007, he was elected for New Century Talent of Chinese Education Ministry and Shanghai Dawn Scholar.

Jia Hou received his B.S. degree in communication engineering from Wuhan University of Technology in 2000, China, and the M.S. and Ph.D. degrees in information communication from Chonbuk National University, Korea, in 2002 and 2005, respectively. He was the post-doctoral research fellow and the invited professor in Chonbuk National University, Korea, from April 2005 to April 2007.

And now he is the associate professor in Soochow University, Suzhou, China. His main research interests are Signal Processing, sequences, error coding and space time signal processing. He has published more than sixty research papers in international journals and conferences proceedings.



Junqiang Fu received his B.S. degree in communication engineering from Tianjin University in 2013, China. He is currently pursuing the Master's degree at Beijing University of Posts and Telecommunications (BUPT). His research interests include modulation recognition and statistical signal processing in time-varying channels.

Chenglin Zhao received the Bachelor's degree in radio-technology from Tianjin University in 1986, and the Master's degree in circuits and systems from Beijing University of Posts and Telecommunications (BUPT) in 1993, and the Ph.D. degree in communication and information system from Beijing University of Posts and Telecommunications, in 1997. At present, he serves as a Professor in Beijing University of Posts and Telecommunications, Beijing, China. His research is focused on emerging technologies of short-range wireless communication, cognitive radios, 60GHz millimeter-wave communications.

Arumugam Nallanathan (S'97–M'00–SM'05) is a Professor of Wireless Communications in the Department of Informatics at King's College London (University of London). He served as the Head of Graduate Studies in the School of Natural and Mathematical Sciences at King's College London, 2011/12. He was an Assistant Professor in the Department of Electrical and Computer Engineering, National University of Singapore from August 2000 to December 2007. His research interests include 5G Technologies, Millimeter wave communications, Cognitive Radio and Relay Networks. In these areas, he co-authored nearly 250 papers. He is a co-recipient of the Best Paper Award presented at the 2007 IEEE International Conference on Ultra-Wideband (ICUWB2007). He is a Distinguished Lecturer of IEEE Vehicular Technology Society.

He is an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS and the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He was an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2006–2011), the IEEE WIRELESS COMMUNICATIONS LETTERS and the IEEE SIGNAL PROCESSING LETTERS. He served as the Chair for the Signal Processing and Communication Electronics Technical Committee of IEEE Communications Society, Technical Program Co-Chair (MAC track) for IEEE WCNC 2014, Co-Chair for the IEEE GLOBECOM 2013 (Communications Theory Symposium), Co-Chair for the IEEE ICC 2012 (Signal Processing for Communications Symposium), Co-Chair for the IEEE GLOBECOM 2011 (Signal Processing for Communications Symposium), Technical Program Co-Chair for the IEEE International Conference on UWB 2011 (IEEE ICUWB 2011), Co-Chair for the IEEE ICC 2009 (Wireless Communications Symposium), Co-chair for the IEEE GLOBECOM 2008 (Signal Processing for Communications Symposium) and General Track Chair for IEEE VTC 2008. He received the IEEE Communications Society SPCE outstanding service award 2012 and IEEE Communications Society RCC outstanding service award 2014.