

# Modulation Classification via Subspace Detection in MIMO Systems

Hadi Srieddeen, *Student Member, IEEE*, Mohammad M. Mansour, *Senior Member, IEEE*, and Ali Chehab

## Abstract

The problem of efficient modulation classification (MC) in multiple-input multiple-output (MIMO) systems is considered. Per-layer likelihood-based MC is proposed by employing subspace decomposition to partially decouple the transmitted streams. When detecting the modulation type of the stream of interest, a dense constellation is assumed on all remaining streams. The proposed classifier outperforms existing MC schemes at a lower complexity cost, and can be efficiently implemented in the context of joint MC and subspace data detection.

## Index Terms

MIMO, adaptive modulation, modulation classification, subspace detection.

## I. INTRODUCTION

Modulation classification is the task of recognizing the modulation type (MT) employed at the transmitter of a detected signal, which is required for various military and civilian applications. In particular, cognitive radio with adaptive MTs [1] is a promising future application of MC. In such scenario, the transmitter dynamically adjusts the data rate by switching the modulation order depending on channel conditions. By employing automatic/blind MC at the receiver, the communication overhead can be reduced.

MC techniques are of two types: feature-based and likelihood-based [2]. While adding more antennas in MIMO technology [3] increases spectral efficiency, the complexity of optimal likelihood-based MC [4]–[6] increases exponentially with this addition, especially when each layer is allowed to have a different MT. An alternative low complexity sub-optimal feature-based MC scheme for MIMO systems exploits the higher-order cyclic cumulants (CCs) of the baseband intercepted signal [7]. Moreover, without perfect channel state information (CSI), independent

component analysis has been used [8] to blindly estimate the channel in conjunction with either likelihood-based or feature-based MC.

In this letter, we propose near-optimal low-complexity likelihood-based MC for MIMO systems. We first decompose the channel matrix employing subspace decomposition, and then detect the MT on the partially decoupled stream of interest using a modified likelihood metric. A joint MC and subspace data detection receiver is also proposed. Regarding notations, bold upper case, bold lower case, and lower case letters correspond to matrices, vectors, and scalars, respectively. Scalar norm, vector norm, conjugate transpose, and inverse are represented by  $|\cdot|$ ,  $\|\cdot\|$ ,  $(\cdot)^*$ , and  $(\cdot)^{-1}$ , respectively.

## II. SYSTEM MODEL

We consider, a spatial multiplexing MIMO system with  $N$  transmit, and for simplicity  $M = N$  receive antennas:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z} \quad (1)$$

with  $\mathbf{y} = [y_1 y_2 \dots y_M]^T \in \mathcal{C}^{M \times 1}$  being the received complex vector,  $\mathbf{H} = [\mathbf{h}_1 \mathbf{h}_2 \dots \mathbf{h}_N] \in \mathcal{C}^{M \times N}$  the complex channel matrix,  $\mathbf{x} = [x_1 x_2 \dots x_N]^T \in \mathcal{C}^{N \times 1}$  the transmitted symbol vector, and  $\mathbf{z} \in \mathcal{C}^{M \times 1}$  the complex additive white Gaussian noise vector with zero mean and variance  $\sigma^2$  ( $\mathbb{E}[\mathbf{z}\mathbf{z}^*] = \sigma^2 \mathbf{I}_N$ ). Each symbol  $x_n$  belongs to a normalized complex constellation  $\mathcal{X}_n$  of size  $Q_n = 2^{q_n}$ , thus  $\mathbf{x} \in \bar{\mathcal{X}} = \mathcal{X}_1 \times \dots \times \mathcal{X}_N$  and  $\mathbb{E}[x_n^* x_n] = 1$ . Consequently, the signal to noise ratio (SNR) is defined in terms of the noise variance as  $\text{SNR} = (N/\sigma^2)$ . The bit representation of symbol  $x_n$  is a coded bit-interleaved sequence  $\mathbf{b}_n = (b_{n,1}, b_{n,2}, \dots, b_{n,q_n})$ .

In a MIMO system that supports non-uniform MTs, each of the  $N$  transmitted symbols is assumed to be drawn from one of  $S$  possible MTs, with equal probability. We develop MC schemes to estimate the MT per-layer, using the received signal  $\mathbf{y}$  and assuming perfect CSI.

## III. LIKELIHOOD-BASED MC

The optimal likelihood-based MC scheme decides on the MT that has the maximum likelihood within multiple hypotheses. Bayesian hypothesis testing is performed on the  $S^N$  possible hypotheses, corresponding to  $\bar{\mathcal{X}}_j = \mathcal{X}_{j,1} \times \dots \times \mathcal{X}_{j,N}$  finite lattices ( $j \in \{1, \dots, S^N\}$ ), with likelihoods:

$$P(\mathbf{y}; \bar{\mathcal{X}}_j) = \sum_{\mathbf{x} \in \bar{\mathcal{X}}_j} P(\mathbf{y}|\mathbf{x})P(\mathbf{x}) \quad (2)$$

where  $P$  is the probability density function. Under statistical independence between the components of  $\mathbf{x}$ , and assuming uniform priors,  $P(x_n) = 1/|\mathcal{X}_n|$ , where  $|\cdot|$  denotes the constellation set cardinality, the decision metric is derived as:

$$\eta = \arg \max_{j \in \{1, \dots, S^N\}} \sum_{\mathbf{x} \in \mathcal{X}_j} P(\mathbf{y}|\mathbf{x}) \frac{1}{|\mathcal{X}_{j,1}|} \times \dots \times \frac{1}{|\mathcal{X}_{j,N}|} \quad (3)$$

Noting that  $P(\mathbf{y}|\mathbf{x}) = \frac{1}{(\pi\sigma^2)^M} \exp(-\frac{1}{\sigma^2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2)$ , and neglecting the term  $\frac{1}{(\pi\sigma^2)^M}$  which is assumed fixed over hypotheses, the resultant Log-MAP decision metric is:

$$\eta_L = \arg \max_{j \in \{1, \dots, S^N\}} \left( \log \frac{1}{|\mathcal{X}_{j,1}|} + \dots + \log \frac{1}{|\mathcal{X}_{j,N}|} + \log \sum_{\mathbf{x} \in \mathcal{X}_j} \exp \left( -\frac{1}{\sigma^2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 \right) \right) \quad (4)$$

which is the average likelihood ratio test (ALRT) solution.

Solving equation (4) is computationally intensive, because for each  $j$  we have to calculate  $|\mathcal{X}_{j,1}| \times \dots \times |\mathcal{X}_{j,N}|$  exponential terms. However, one of these terms is dominant and corresponds to the scaled maximum likelihood (ML) distance:

$$d_{ML,j} = \min_{\mathbf{x} \in \mathcal{X}_j} \frac{1}{\sigma^2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 \quad (5)$$

Using the approximation  $\log \sum_r \exp(a_r) \approx \max_r \{a_r\}$ , we get:

$$\eta_M = \arg \max_{j \in \{1, \dots, S^N\}} \left( \log \frac{1}{|\mathcal{X}_{j,1}|} + \dots + \log \frac{1}{|\mathcal{X}_{j,N}|} - d_{ML,j} \right) \quad (6)$$

which is the near-optimal Max-Log-MAP classifier.

While Max-Log-MAP eliminates exponential operations, the number of Euclidean distance computations per hypothesis remains exponential, and computing the likelihood functions  $S^N$  times is exhaustive. An alternative approach is required, that separates the transmitted signals for individual treatment, which results in only  $|\mathcal{X}_{j,n}|$  distance computations per layer  $n$  and hypothesis  $j \in \{1, \dots, S\}$ . This is achieved by the per-layer sub-optimal ALRT solution. With perfect CSI at the receiver, the sub-optimal ALRT classifier finds the zero-forcing (ZF) equalized output  $\hat{\mathbf{y}}_{ZF} = (\mathbf{H}^* \mathbf{H})^{-1} \mathbf{H}^* \mathbf{y}$ , computes the scaled noise variance  $\sigma_{ZF}^2 = (\mathbf{h}_n^* \mathbf{h}_n)^{-1} \sigma^2$ , and generates the likelihood function per layer  $n$  as follows:

$$\eta_S = \arg \max_{j \in \{1, \dots, S\}} \left( \log \frac{1}{|\mathcal{X}_{j,n}|} + \log \sum_{x_n \in \mathcal{X}_{j,n}} \exp \left( -\frac{1}{\sigma_{ZF}^2} |\hat{y}_{ZF,n} - x_n|^2 \right) \right) \quad (7)$$

We seek a classifier that decouples the layers while maintaining distance metrics that are close to that of Log-MAP.

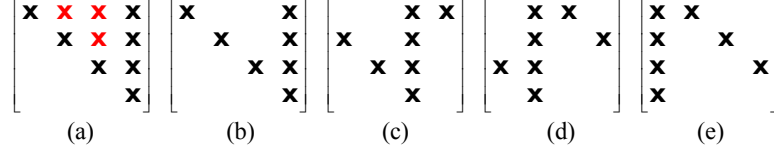


Fig. 1. 4x4 channel matrix structures

#### IV. PROPOSED MC SCHEME

##### A. Subspace Decomposition

While the standard QR decomposition (QRD) decomposes the channel matrix  $\mathbf{H}$  into an upper triangular matrix (UTM)  $\mathbf{R} = [u_{pq}] \in \mathcal{C}^{N \times N}$  with  $u_{pp} \in \mathcal{R}^+$ , and a unitary matrix  $\mathbf{Q}$ , the so-called WR decomposition (WRD) scheme punctures in addition the red-marked entries above the diagonal in Fig. 1(a). We aim at transforming  $\mathbf{H}$  into a punctured UTM  $\mathbf{R}$  as shown in Fig. 1(b), through a matrix  $\mathbf{W} = [\mathbf{w}_1 \mathbf{w}_2 \dots \mathbf{w}_N] \in \mathcal{C}^{N \times N}$ , such that  $\mathbf{W}^* \mathbf{H} = \mathbf{R}$ . We assume  $\mathbf{H} = [\mathbf{h}_1 \mathbf{h}_2 \dots \mathbf{h}_N]$  to have a full column rank, and impose the condition on the column vectors of  $\mathbf{W}$  to have unit length, i.e.,  $\mathbf{w}_n^* \mathbf{w}_n = 1$  for  $n = 1, \dots, N$ , so that the transformed noise vector will maintain an unaltered covariance matrix ( $\mathbb{E}[\mathbf{W}^* \mathbf{z} \mathbf{z}^* \mathbf{W}] = \sigma^2 \mathbf{I}_N$ ).

The brute force approach for computing  $\mathbf{W}$  involves extensive matrix inversions, which is computationally intensive and prone to numerical error. An alternative approach [9] consists of QRD followed by elementary matrix operations. Let  $\mathbf{H}$  be QR-decomposed such that  $\mathbf{Q}_1^* \mathbf{H} = \mathbf{R}_1$ . Consider row  $1 < n \leq N$  of  $\mathbf{R}_1$ , and assume the  $m$ th entry  $r_{nm}$ ,  $m > n$ , is to be nulled. We have  $\mathbf{q}_n^* \mathbf{h}_m = r_{nm} \in \mathcal{C}$  and  $\mathbf{q}_m^* \mathbf{h}_m = r_{mm} \in \mathcal{R}^+$ , from which it follows that  $(\mathbf{q}_n^* - \mathbf{q}_m^* \frac{r_{nm}}{r_{mm}}) \mathbf{h}_m = 0$ . Therefore:

$$\mathbf{q}_n = \mathbf{q}_n - \mathbf{q}_m r_{nm}^* / r_{mm} \quad (8)$$

$$r_{nj} = r_{nj} - r_{mj} r_{nm} / r_{mm}, \quad \text{for } j = m, \dots, N \quad (9)$$

puncture the required entry and update  $\mathbf{Q}_1$  accordingly. Finally, the non-zero entries in row  $n$  of  $\mathbf{R}_1$  are updated, and  $\mathbf{q}_n$  is normalized to have unit length:

$$r_{nj} = r_{nj} / \|\mathbf{q}_n\|, \quad \text{for } j = n, \dots, N \quad (10)$$

$$\mathbf{q}_n = \mathbf{q}_n / \|\mathbf{q}_n\| \quad (11)$$

Repeating this on required entries (bottom to top, right to left), the resulting  $\mathbf{Q}_1$  is  $\mathbf{W}$ , and  $\mathbf{R}_1$  is the desired punctured  $\mathbf{R}$ .

### B. Proposed Likelihood-Based MC

To generate the likelihood functions on all layers, the  $N$  streams are decoupled, one at a time, by cyclically shifting the columns of  $\mathbf{H}$  and generating the punctured UTMs, as shown in Fig. 1(b-e). Alternatively, a minimal swapping operation can put the layer of interest at the rightmost column location, and hence a decomposition as shown in Fig. 1(b) will always follow. We represent this swapping operation by a permutation:

$$\pi^{(n)}(i) = \begin{cases} N & \text{if } i = n \\ n & \text{if } i = N \\ i & \text{otherwise} \end{cases} \quad (12)$$

for  $i = 1, \dots, N$  and at a layer of interest  $n$ . Each permuted  $\mathbf{H}^{(n)}$  is then WR-decomposed into  $\mathbf{W}^{(n)}$  and  $\mathbf{R}^{(n)}$ .

We first partition  $\tilde{\mathbf{y}}^{(n)}$ ,  $\mathbf{R}^{(n)}$ , and permuted  $\mathbf{x}$  as:

$$\tilde{\mathbf{y}}^{(n)} = \begin{bmatrix} \tilde{\mathbf{y}}_1^{(n)} \\ \tilde{y}_2^{(n)} \end{bmatrix}, \quad \mathbf{R}^{(n)} = \begin{bmatrix} \mathbf{A}^{(n)} & \mathbf{b}^{(n)} \\ 0 & c^{(n)} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \tilde{x}_2 \end{bmatrix} \quad (13)$$

where  $\tilde{\mathbf{y}}_1^{(n)} \in \mathcal{C}^{(N-1) \times 1}$ ,  $\tilde{y}_2^{(n)} \in \mathcal{C}^{1 \times 1}$ ,  $\mathbf{A}^{(n)} \in \mathcal{R}^{(N-1) \times (N-1)}$ ,  $\mathbf{b}^{(n)} \in \mathcal{C}^{(N-1) \times 1}$ ,  $c^{(n)} \in \mathcal{R}^{1 \times 1}$ ,  $\mathbf{x}_1 \in \mathcal{X}^{N-1}$ , and  $\tilde{x}_2 \in \mathcal{X}_n$ . Then, the modified distance metric is expressed in terms of  $\tilde{x}_2$  as:

$$\|\tilde{\mathbf{y}}^{(n)} - \mathbf{R}^{(n)} \mathbf{x}\|^2 = \left| \tilde{y}_2^{(n)} - c^{(n)} \tilde{x}_2 \right|^2 + \left\| \tilde{\mathbf{y}}_1^{(n)} - \mathbf{A}^{(n)} \hat{\mathbf{x}}_1 - \mathbf{b}^{(n)} \tilde{x}_2 \right\|^2 \quad (14)$$

$$\hat{\mathbf{x}}_1 = \lfloor (\tilde{\mathbf{y}}_1^{(n)} - \mathbf{b}^{(n)} \tilde{x}_2) / \mathbf{A}^{(n)} \rfloor_{\mathcal{X}^{N-1}} \quad (15)$$

where  $\lfloor \alpha \rfloor_{\mathcal{X}_r} \triangleq \arg \min_{x_r \in \mathcal{X}_r} |\alpha - x_r|$  is the slicing operator. Since  $\mathbf{A}^{(n)}$  is a diagonal matrix, slicing is applied to the individual elements of the vector  $\tilde{\mathbf{y}}_1^{(n)}$  in parallel.

Accumulating  $T$  observations before deciding on a winning hypothesis, the proposed likelihood functions at layer  $n$  are:

$$\hat{\eta}_{\text{L}} = \arg \max_{j \in \{1, \dots, S\}} \sum_{t=1}^T \left( \log \frac{1}{|\mathcal{X}_{j,n}|} + \log \sum_{\tilde{x}_2 \in \mathcal{X}_{j,n}} \exp \left( -\frac{1}{\sigma^2} \|\tilde{\mathbf{y}}^{(n)} - \mathbf{R}^{(n)} \mathbf{x}\|^2 \right) \right) \quad (16)$$

$$\hat{\eta}_{\text{M}} = \arg \max_{j \in \{1, \dots, S\}} \sum_{t=1}^T \left( \log \frac{1}{|\mathcal{X}_{j,n}|} - \hat{d}_{\text{ML},j} \right) \quad (17)$$

TABLE I  
COMPUTATIONAL COMPLEXITIES OF MC SCHEMES

Approach	Euc. Dist.	Exp.	Log.
Log-MAP (ALRT)	$S^N \times  \mathcal{X}_{\max} ^N$	$S^N \times  \mathcal{X}_{\max} ^N$	$S^N$
Max-Log-MAP	$S^N \times  \mathcal{X}_{\max} ^N$	0	0
Sub-optimal ALRT	$N \times S \times  \mathcal{X}_{\max} $	$N \times S \times  \mathcal{X}_{\max} $	$S$
Subspace-Log-MAP	$N \times S \times  \mathcal{X}_{\max} $	$N \times S \times  \mathcal{X}_{\max} $	$S$
Subspace-Max-Log-MAP	$N \times S \times  \mathcal{X}_{\max} $	0	0

$$\hat{d}_{\text{ML},j} = \min_{\tilde{x}_2 \in \mathcal{X}_{j,n}} \frac{1}{\sigma^2} \|\tilde{\mathbf{y}}^{(n)} - \mathbf{R}^{(n)} \mathbf{x}\|^2 \quad (18)$$

Note that optimal slicing in equation (15) requires knowledge of MTs on all remaining layers, which is infeasible in an independent per-layer scheme. Therefore, we propose to do slicing assuming dense constellations,  $\mathcal{X} = 1024\text{-QAM}$  for example. The idea of slicing over a dense constellation comes from the work in [6] on multiuser MIMO (MU-MIMO) detection, where it has been shown that near-optimal data detection can be achieved while assuming interferers to have high order MTs, which captures the geometry of constellations while minimizing errors. Note that 1024-QAM was not considered in [6] because of the entailed complexity of ML detection. This is not an issue in our case, since subspace detection only employs the dense constellations in slicing operations.

Table I compares the upper bound on computational complexities of studied classifiers, in terms of the number of Euclidean distance computations, as well as exponential and logarithmic operations, where  $\mathcal{X}_{\max}$  is the largest possible MT. Note that the table does not account for the less significant preprocessing computations (ZF equalization, QRD/WRD) that can be computed once for a large number of observations when the channel variation is slow. While computations in optimal ALRT are exponential in the number of transmit antennas, they are linear in the proposed subspace-based classifiers and sub-optimal ALRT solution (the latter is less complex since its distance computations are one dimensional).

## V. JOINT MC AND SUBSPACE DETECTION

In order to generate the log-likelihood ratio (LLR) of the  $k$ th bit of the  $n$ th symbol  $b_{n,k}$  via subspace detection, we compute two distance metrics defined as:

$$\mathbf{u}_{n,k} = \arg \min_{x_n \in \mathcal{X}_{n,k}^0} \|\tilde{\mathbf{y}}^{(n)} - \mathbf{R}^{(n)} \mathbf{x}\|^2, \mathbf{v}_{n,k} = \arg \min_{x_n \in \mathcal{X}_{n,k}^1} \|\tilde{\mathbf{y}}^{(n)} - \mathbf{R}^{(n)} \mathbf{x}\|^2 \quad (19)$$

where  $k = 1, \dots, q_n$ , the sets  $\mathcal{X}_{n,k}^{(0)} = x_n \in \mathcal{X}_n : b_{n,k} = 0$  and  $\mathcal{X}_{n,k}^{(1)} = x_n \in \mathcal{X}_n : b_{n,k} = 1$  correspond to subsets of symbol vectors in  $\mathcal{X}_n$ , having in the corresponding  $k$ th bit of the  $n$ th symbol a value of 0 and 1, respectively, and the distance metrics are expanded as in equation (14). The unscaled LLRs are then calculated as:

$$\Lambda_{n,k} = \|\tilde{\mathbf{y}}^{(n)} - \mathbf{R}^{(n)} \mathbf{u}_{n,k}\|^2 - \|\tilde{\mathbf{y}}^{(n)} - \mathbf{R}^{(n)} \mathbf{v}_{n,k}\|^2 \quad (20)$$

If  $\mathbf{R}$  is unpunctured, the parallel slicing operation of equation (15) can not be executed, and successive interference cancellation (SIC) is applied to obtain  $\hat{\mathbf{x}}_1$ . Expanding the distances in equation (20) via SIC results in a special subspace detector called layered orthogonal lattice detector (LORD) [10], and expanding the distances in equations (16) and (17) accordingly results in LORD-Log-MAP and LORD-Max-Log-MAP classifiers, respectively.

Be it detection or MC, while independently processing a layer of interest, the MTs on the remaining layers are unknown, and parallel slicing or SIC is conducted assuming 1024-QAM. This means that the distance metrics computed for data detection are identical to those computed in equations (16) and (17) for the winning hypothesis, and thus combining MC and detection results in a minimal MC overhead.

The joint MC and detection setup is summarized in algorithm 1 and architecturally illustrated in Fig. 2. For  $T$  observations, the detection routine is called  $T$  times for all hypotheses, and the resulting distance metrics are stored in memory. Concurrently, the likelihood of each hypothesis is computed. Eventually, the metrics corresponding to the winning hypothesis are retrieved for LLR processing. The receiver can run in this joint mode for a sufficient number of observations and then switch back to regular data detection. Moreover, since the operations on different layers are independent, the proposed algorithm can be parallelized on multiple processing units. Finally, if the system had more receive antennas ( $M > N$ ), the “thin” form of the QR decomposition for tall matrices can be used, and other modifications immediately follow.

---

**Algorithm 1** Proposed per-layer joint MC and detection
 

---

- 1: Swap the column of interest  $n$  with column  $N$  in  $\mathbf{H}$  as in equation (12)
  - 2: Decompose the channel matrix as in equation (13)
  - 3: Calculate the distance metrics for all hypotheses as in equation (14) while assuming the MTs on the remaining layers to be 1024-QAM
  - 4: Calculate the classifier likelihood function as in equations (16) and (17)
  - 5: Repeat steps 3 and 4 for  $T$  observations, accumulate likelihoods, and decide on the winning hypothesis
  - 6: Forward the distance metrics that correspond to the winning hypothesis for bit LLR generation as in equation (20)
- 

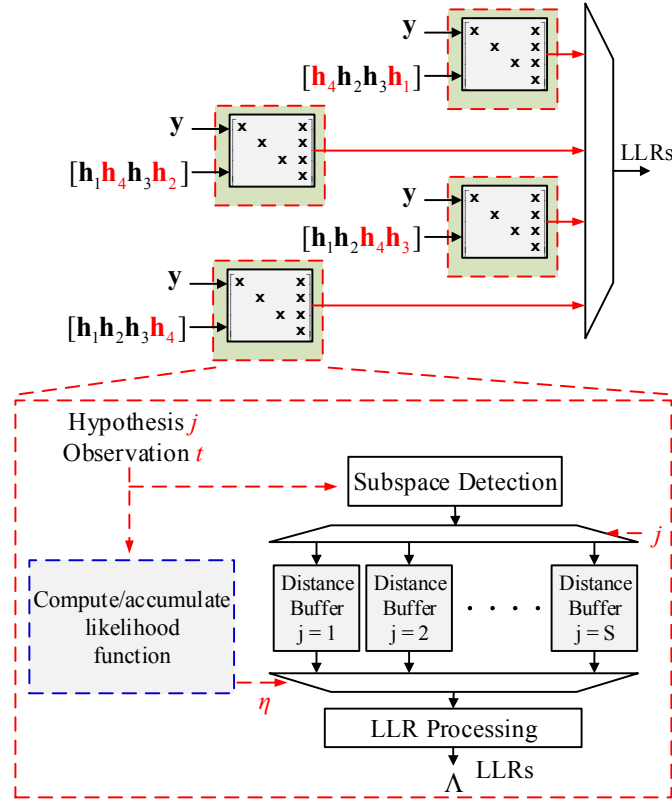


Fig. 2. Joint MC and subspace detection architecture

## VI. SIMULATION RESULTS

Several MC and detection schemes were simulated in the context of  $4 \times 4$  MIMO. We considered five hypotheses of MTs per layer, varying with equal probability on every new frame, which are  $\phi$ ,



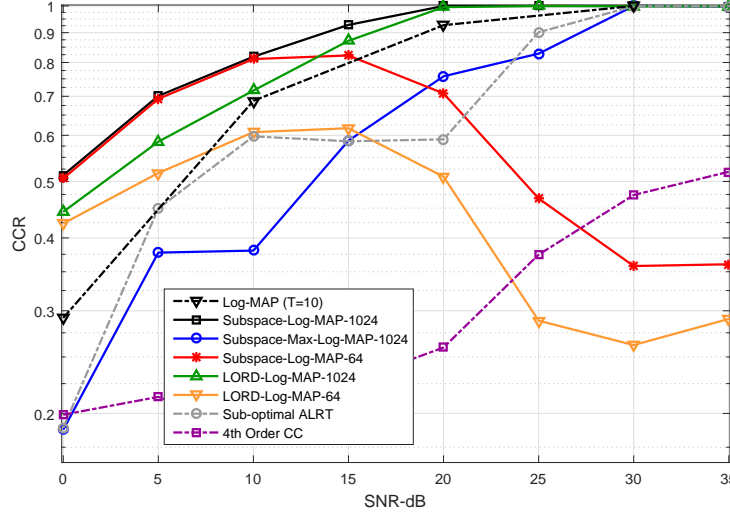


Fig. 3. CCR performance - uncorrelated channels

QPSK, 16-QAM, 64-QAM and 256-QAM, with  $\phi$  representing a constellation having one entry of zero power, corresponding to the case when the transmitting antenna is silent. Note that an all-QAM set of hypotheses that only differs by modulation order is hard to classify, but is more likely to occur in future standards. The winning hypothesis was decided after accumulating  $T = 1000$  observations. Turbo coding was used, with a code rate of  $1/2$  and 8 decoding iterations. Moreover, in addition to the regular channel  $\mathbf{H}$ , we considered a correlated channel  $\mathbf{H}_c = \mathbf{R}_r^{1/2} \mathbf{H} \mathbf{R}_t^{1/2}$ , where  $\mathbf{R}_t$  and  $\mathbf{R}_r$  are the transmit and receive antenna correlation matrices, respectively, with a correlation factor of 0.3.

The proposed classifiers are compared in terms of correct classification ratio (CCR). Fig. 3 shows that for uncorrelated channels, the best performance is achieved by Subspace-Log-MAP and LORD-Log-MAP classifiers, when the MTs on the remaining layers are assumed to be 1024-QAM. The Subspace-Max-Log-MAP and sub-optimal ALRT classifiers lag behind, but are also capable of achieving unity CCR at high SNR. However, assuming 64-QAMs instead of 1024-QAMs resulted in bad classification performance for both subspace and LORD based classifiers. The exhaustive Log-MAP classifier with  $T = 10$  observations only was added as a reference, and the much less complex proposed approaches outperformed it. Also, the reference feature based (4th order CCs) classifier performed very bad with an all-QAM set of hypotheses. Fig. 4 then shows that for highly correlated channels, only Subspace-Log-MAP performs well

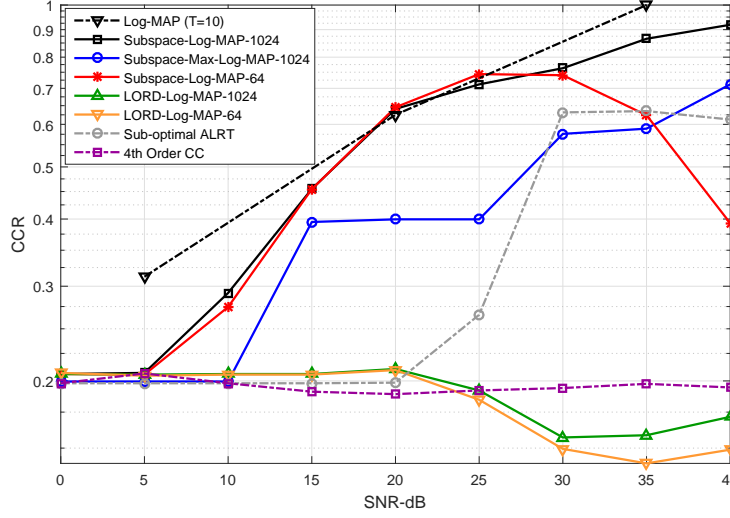


Fig. 4. CCR performance - correlated channels

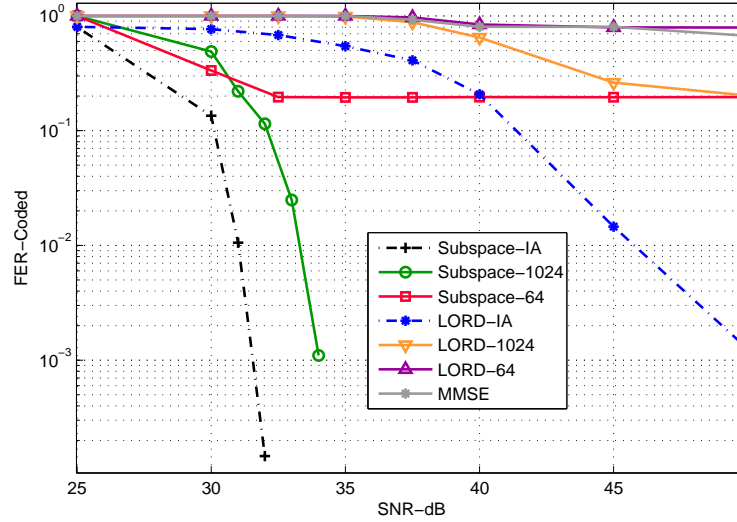


Fig. 5. FER performance - correlated channels - 64-QAM on layer of interest

at high SNR, approaching the upper Log-MAP bound.

The corresponding coded frame error rate (FER) performance of the proposed detectors with high channel correlation is shown in Fig. 5. The detectors were simulated assuming the layer of interest to use 64-QAM (following successful per-layer MC), while the MTs at the remaining layers were unknown, and randomly hopping over possible hypotheses. Both LORD and subspace detectors were tested, assuming the remaining MTs to be 1024-QAM or 64-QAM.

These detectors were compared to the regular MT-aware LORD and subspace detectors that have perfect knowledge of MTs on all layers. While regular subspace detection beats LORD by more than 10 dB, only assuming 1024-QAM in conjunction with subspace detection was able to achieve near MT-aware performance. This declares the subspace-based classifiers winners in the context of joint MC and detection. Finally, not shown in this letter for lack of space, are the FER plots with uncorrelated channels, where all schemes achieved near-MT-aware performance, and the CCR plots with imperfect CSI, that show remarkable performance deterioration (mitigating channel estimation errors is beyond the scope of this work).

## VII. CONCLUSION

Low-complexity per-layer MC schemes have been proposed for MIMO systems, based on subspace decomposition. It has been shown that assuming the modulation type on all layers except the layer of interest to be a dense constellation results in good classification performance. This assumption has been proved to have a negligible performance degradation cost in subspace data detection, which made fully parallelizable efficient joint MC and data detection feasible.

## REFERENCES

- [1] H. Arslan, *Cognitive radio, software defined radio, and adaptive wireless systems*. Springer, 2007.
- [2] O. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: classical approaches and new trends," *Communications, IET*, vol. 1, no. 2, pp. 137–156, Apr. 2007.
- [3] A. Paulraj, R. Nabar, and D. Gore, *Introduction to space-time wireless communications*. Cambridge University Press, 2003.
- [4] Y. Eldemerdash, O. Dobre, and M. Oner, "Signal identification for multiple-antenna wireless systems: Achievements and challenges," *IEEE Communications Surveys Tutorials*, vol. 18, no. 3, pp. 1524–1551, 2016.
- [5] V. Choqueuse, S. Azou, K. Yao, L. Collin, and G. Burel, "Blind modulation recognition for MIMO systems," *MTA Review*, vol. 19, no. 2, pp. 183–196, 2009.
- [6] H. Srieddeen, M. Mansour, L. Jalloul, and A. Chehab, "Efficient near optimal joint modulation classification and detection for MU-MIMO systems," in *ICASSP*, 2016, pp. 3706–3710.
- [7] M. Muhlhaus, M. Oner, O. Dobre, H. Jakel, and F. Jondral, "A novel algorithm for MIMO signal classification using higher-order cumulants," in *Radio and Wireless Symposium (RWS), 2013 IEEE*, 2013, pp. 7–9.
- [8] M. Muhlhaus, M. Oner, O. Dobre, and F. Jondral, "A low complexity modulation classification algorithm for MIMO systems," *Communications Letters, IEEE*, vol. 17, no. 10, pp. 1881–1884, Oct. 2013.
- [9] M. Mansour, "A near-ML MIMO subspace detection algorithm," *Signal Processing Letters, IEEE*, vol. 22, no. 4, pp. 408–412, Apr. 2015.
- [10] M. Siti and M. Fitz, "A novel soft-output layered orthogonal lattice detector for multiple antenna communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 4, 2006, pp. 1686–1691.