Hierarchical Coherent and Non-coherent Communication

Kareem M.Attiah*, Karim G. Seddik† and Ramy H. Gohary‡
*Department of Electrical Engineering, Alexandria University, Alexandria 21544, Egypt
†Electronics Engineering Department, American University in Cairo, AUC Avenue, New Cairo 11835, Egypt
‡Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada

***Abstract*—In this paper, we propose a method for simultaneous communication of coarse and detailed information over three types of layers: two nested layers to be communicated non-coherently using Grassmannian constellations and an additional layer to be communicated coherently using unitary constellations. The layered architecture gives rise to four classes of receivers: coherent and non-coherent receivers, each operating in one of two distinct SNR regions. An operational bottleneck of this architecture is the detection complexity of these layers. To overcome this difficulty, four sequential detectors are developed. These detectors enable somewhat comparable, and in some cases identical, performance to that of their optimal maximum likelihood counterparts with significantly less computational cost.**

I. Introduction

Conventional communications typically use single-layer transmission to communicate information over the channel. In this framework, transmitted segments of information are considered to be equally valuable for the prospective destination. Contrary to single-layer transmission are the multi-layer ones. In multi-layer systems, various receivers may be able to reconstruct some segments of the transmitted information, but not the others, depending on their channel conditions, their received signal-to-noise ratio (SNR) and their computational power. For instance, a particular receiver with favourable channel conditions might be able to recover all the information segments, whereas another receiver with less favourable channel conditions will only be able to recover the more basic information segments.

Multi-layered signalling has many applications, e.g., in military [1] and multimedia [2] communications. Most of the work in this area focuses on receivers operating over distinct SNR regions. However, other scenarios may arise in which some receivers have more information about the channel than others. In an extreme case, some receivers may experience low mobility and therefore may have perfect channel state information (CSI), whereas other receivers may experience high mobility and therefore may have no access to CSI. Focusing on this extreme case, the framework in [3] proposes a signalling scheme that is conducive to effective communication in multiple-input multiple-output (MIMO) systems. In this scheme basic low-resolution (LR) information is communicated in the subspace spanned by the transmitted signal matrix and detailed high-resolution (HR) information is communicated in the basis that spans this subspace. The LR information can be recovered non-coherently, i.e., without invoking CSI, but the HR information must be recovered coherently and therefore requires reliable CSI to be available at the receiver.

To realize the multi-layered signalling scheme in [3], the transmitted signals are restricted to be in the form of tall unitary matrices. The subspace spanned by one such matrix can be represented by a point on the compact Grassmann manifold [4]. Such a point is invariant under right multiplication by square unitary matrices. Using this feature, it can be seen that the role of the right multiplication by a square unitary matrix is to specify the basis that spans the subspaces spanned by the tall unitary component. Interestingly, using such a multiplication not only enables concurrent transmission of LR and HR information, but also does that without incurring additional power consumption. In other words, the HR information will be communicated with essentially zero additional power. Furthermore, multiplying points of the Grassmannian constellation by square unitary matrices does not affect the distance properties and hence does not compromise the performance yielded by these constellations.

Building on the work in [3], herein we introduce a layered structure within the Grassmannian constellation. In particular, the Grassmannian constellation will be conceived as the Cartesian product of a ‘parent’ constellation and a ‘child’ one. This structure will enable non-coherent receivers operating at lower SNRs to have access only to the LR information encoded in the parent constellation, whereas non-coherent receivers operating at higher SNRs will have access to HR information encoded in both the parent and child constellations. The composite Grassmannian point comprised of both a parent component and a child one are then right multiplied by point from a square unitary constellation to specify the basis of the subspace spanned by the composite point. Similar to [3], these bases will be used to communicate HR information to coherent receivers.

The contributions in this paper are summarized in the following points.

- We propose a three-layer technique consisting of a coherent and multi-layer non-coherent components, thereby enriching the flexibility with which HR information can be communicated.
- We derive conditional distributions of the received signal and utilize these expressions to derive the respective maximum-likelihood (ML) decision criteria.
- We develop simplified detectors that yield close-to-optimal performance but at a significantly less computational cost.

II. Preliminaries and System Model

A. Preliminaries and Non-coherent Communication

The unitary group, denoted by $\mathbb{U}_M$, is the set of $M \times M$ square unitary matrices. This set is closed under matrix multiplication, i.e. $\mathbf{A}_1\mathbf{A}_2 \in \mathbb{U}_M$, for all $\mathbf{A}_1, \mathbf{A}_2 \in \mathbb{U}_M$. Elements of this set are of M^2 real dimensions.

For $T > M$, the set of $T \times M$ complex matrices with orthonormal columns is known as the **Stiefel Manifold**, $\mathbb{S}_{T,M}(\mathbb{C}) = \{\mathbf{P} \in \mathbb{C}^{T\times M} | \mathbf{P}^\dagger \mathbf{P} = \mathbf{I}_M\}$. For two matrices $\mathbf{P}_1, \mathbf{P}_2 \in \mathbb{S}_{T,M}(\mathbb{C})$, an equivalence relation can be defined whereby $\mathbf{P}_1$ and $\mathbf{P}_2$ are equivalent if and only if the columns of both matrices span the same subspace, that is $\mathbf{P}_1 = \mathbf{P}_2\mathbf{Z}$, for some $\mathbf{Z} \in \mathbb{U}_M$. The **Grassmann Manifold**, $\mathbb{G}_{T,M}(\mathbb{C})$, is the quotient space $\mathbb{S}_{T,M}(\mathbb{C})/\mathbb{U}_M$ with respect to the equivalence relation. In other words, it can be regarded as the set of M dimensional subspaces in $\mathbb{C}^T$. Elements in $\mathbb{G}_{T,M}(\mathbb{C})$ are invariant under right-multiplication by elements of $\mathbb{U}_M$. This property will be of key importance in subsequent discussion.

The Grassmann manifold is of particular importance in the context of non-coherent signaling over block fading channels in which both the transmitter and receiver do not have access to CSI. The significance follows from the observation that the channel matrix rotates and scales the transmitted signal basis but does not change the subspace it spans. Thus, in non-coherent communication regimes the information to be transmitted is mapped to a set of subspaces, or alternatively Grassmannian points, which must be well-spaced to ensure that the perturbations induced by channel noise have minimal impact on the receiver's decision. This calls for an appropriate distance metric that reflects the performance of non-coherent constellations. A number of good metrics were proposed in the literature to guide the design of non-coherent constellations. In this paper, we adopt the chordal Frobenius distance [5], whereby the distance between two $T \times M$ unitary matrices, $\mathbf{Q}_1$ and $\mathbf{Q}_2$, is given by

$$d^2(\mathbf{Q}_1, \mathbf{Q}_2) = M - \|\mathbf{Q}_1^\dagger \mathbf{Q}_2\|_F^2 = \sum_{i=1}^{M} \sin^2\theta_i, \tag{1}$$

where $\theta_i, i = 1, \ldots, M$ denote the principle angles between the subspaces spanned by $\mathbf{Q}_1$ and $\mathbf{Q}_2$.

B. Channel Model

We consider a multicast scenario in which a single transmitter wishes to send information to multiple receivers over a block Rayleigh fading channel. In this model, the channel coefficients are essentially fixed during T (henceforth referred to as a coherence interval) symbol durations, then assume an entirely independent realization in later coherence intervals [6]. Letting M (N_i) denote the number of transmit (receive) antennas, the $T \times N_i$ matrix observed by the i-th receiver can be expressed as

$$\mathbf{Y}_i = \mathbf{X}\mathbf{H}_i + \sqrt{M/\gamma T}\,\mathbf{W}_i, \tag{2}$$

where $\mathbf{X}$ is the complex $T \times M$ transmitted matrix, γ is the SNR, and $\mathbf{H}_i \in \mathbb{C}^{M\times N_i}$ and $\mathbf{W}_i \in \mathbb{C}^{T\times N_i}$ are i-th receiver's channel and noise matrices, respectively. Elements of both matrices are independent identically distributed (i.i.d.) complex circularly-symmetric Gaussian distributed, $\mathcal{CN}(0, 1)$.

Our goal herein is for the transmitted signal to be structured in such a way that enables all receivers to decode the fundamental portion of the message encoded in the basic layer, and, depending on the availability of the CSI and the SNR, different receivers will have access to improved reconstructions of the message by retrieving additional layers. In [3], it was proposed that the transmitted signal $\mathbf{X}$ has the form

$$\mathbf{X} = \mathbf{Q}\mathbf{A}, \tag{3}$$

where matrix $\mathbf{Q}$ conveys LR information. It is drawn from a finite subset of $\mathbb{G}_{T,M}(\mathbb{C})$ and can be non-coherently decoded by all receivers. In contrast, the matrix $\mathbf{A}$ conveys HR information. It is drawn from a finite subset of $\mathbb{U}_M$ and must be decoded coherently. Since $\mathbf{X}$ and $\mathbf{Q}$ span the same subspace, a non-coherent receiver will be able to recover the LR information encoded in that subspace. In contrast, a coherent receiver will be able to detect both the subspace and the particular basis specified by $\mathbf{A}$, thereby retrieving both the LR and the HR components of the transmitted message. Herein, we will extends this philosophy by endowing the non-coherent component with a layered structure. The advantage of this setup will be discussed in Sections III and IV.

III. Non-coherent Code Construction

Non-coherent communication at high data rates requires the design of isotropically-distributed Grassmannian constellations [7]. Unfortunately, this task is generally challenging. Design methods for such constellations have been developed in [8], [9]. Although these methods tend to yield constellations with favourable performance, these constellations suffer from several drawbacks. For instance, they do not possess a known structure, which usually renders their storage, generation and detection rather unwieldy. Furthermore, the optimization underlying these designs is rather difficult to solve for medium-size, let alone large-size, constellations. To circumvent this difficulty, a structured approach for designing Grassmannian constellations was developed in [10]. The constellations generated by this approach not only possess a layered structure, but also admit computationally-efficient detection. For completeness, this approach will be discussed next.

Let $\mathcal{C}_L$ be a small-size constellation with points well-spaced, according to the chordal distance, on $\mathbb{G}_{T,M}(\mathbb{C})$. Such a constellation can designed using the method in [11], which was shown to yield near-optimal packings. Suppose now that each point of $\mathcal{C}_L$ is replaced by a cloud of points so that the overall constellation size is $|\mathcal{C}_L||\mathcal{C}_M|$, $|\mathcal{C}_M| > 1$. This idea can be readily extended to a general L-layer construction, but for ease of exposition, we will focus on the case of $L = 2$ layers. The new constellation is formed by means of a basic and an additional *refinement* layer. A simple method for constructing the extended constellation is as follows: Starting at each point

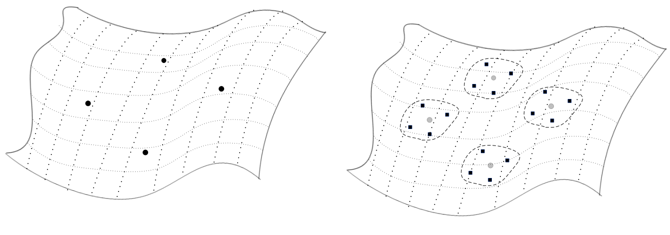

Fig. 1. Pictorial illustration of the proposed multi-layer non-coherent constellation. (Left) Basic constellation points (black circles) on $\mathbb{G}_{T,M}(\mathbf{C})$. (Right) Adding new points in an enclosing region of each basic point. In this case, the new constellation points (black squares) are found by transitioning along $K = 4$ geodesic directions and discarding basic points.

of the basic constellation, find K *children* points by transitioning along K geodesic directions, where $K = |\mathcal{C}_M| - 1$ and include the basic constellation. Alternatively, set $K = |\mathcal{C}_M|$ and discard the basic constellation altogether. In this paper, we adopt the latter approach, cf. Fig. 1.

To obtain Grassmannian constellations with favourable design characteristics, the K geodesic directions ought to be carefully chosen. One way to do so is to choose these directions so that the the chordal distance between children points of a common *parent* (i.e. point in the basic constellation) is maximized. This criterion is derived in [10] and will be discussed later in this section. Another consideration is the distance at which children points lie from their respective parents. If such a distance is chosen to be small, the distance between children associated with a common parent will also be small and the refinement constellation will feature unfavourable distance characteristics. On the other hand, when a large value of this distance is selected, children points will cross over the parent region boundaries thus leading in performance degradation in both layers. In multi-layer applications of interest, where the basic layer is desired to exhibit superior performance, the numerical value of this distance is typically selected to strike a balance between the performance of the refinement and basic layers.

The geodesic of length t emanating from a point $\mathbf{U} = \mathbf{U}(0)$ moving along the direction $\dot{\mathbf{U}}(0) = \mathbf{U}_\perp \mathbf{B}$ for a $(T-M) \times M$ complex matrix $\mathbf{B}$ with SVD $\mathbf{P\Sigma V}^\dagger$ is given by [4]

$$\mathbf{U}(t) = \begin{bmatrix} \mathbf{U} & \mathbf{U}_\perp \end{bmatrix} \begin{bmatrix} \mathbf{V} \cos \mathbf{\Sigma} t \\ \mathbf{P} \sin \mathbf{\Sigma} t \end{bmatrix}, \tag{4}$$

where $\mathbf{U}_\perp$ is a unique representative of the null space of $\mathbf{U}$. In other words, there is a unique matrix $\mathbf{U}_\perp$ corresponding to a given $\mathbf{U}$. Using (4), the extended constellation comprising the basic and refinement constellations can be expressed as

$$\mathcal{C}_{\text{NCOH}} = \left\{ \mathbf{Q} = [\mathbf{U} \quad \mathbf{U}_\perp] \begin{bmatrix} \mathbf{V}_i \cos \mathbf{\Sigma}_i t \\ \mathbf{P}_i \sin \mathbf{\Sigma}_i t \end{bmatrix} \middle| \; \forall \mathbf{U} \in \mathcal{C}_L, i \in \{1, \ldots, |\mathcal{C}_M|\} \right\},$$

Focusing on the important special case of $T = 2M$ and applying the maximum distance criterion for children points of common parent yields [10]

$$\mathcal{C}_{\text{NCOH}} = \left\{ \mathbf{Q} = [\mathbf{U} \; \mathbf{U}_\perp] \begin{bmatrix} \alpha \mathbf{V} \\ \beta \mathbf{I} \end{bmatrix} \middle| \; \forall \mathbf{U} \in \mathcal{C}_L, \mathbf{V} \in \mathcal{C}_M \right\},$$

where $\alpha = \cos t$, $\beta = \sin t$. Using these parameters, it can be seen that the children-parent chordal distance is given by

$$d^2(\mathbf{Q}(\mathbf{U}), \mathbf{U}) = M\beta^2.$$

Finally, the refinement layer $\mathcal{C}_M$, is characterized by the set $\{\mathbf{V}_i\}_{i=1}^{|\mathcal{C}_M|}$, which is obtained by solving the optimization problem, cf. [10].

$$\max_{\{\mathbf{V}_k^\dagger \mathbf{V}_k = \mathbf{I}_M\}_{1 \le k \le K}} \min_{i \neq j} \|\mathbf{V}_i - \mathbf{V}_j\|_F^2.$$

We conclude this section by noting that allowing the non-coherent component to admit the multi-layer structure introduces additional freedom in transmitting information, which can be invoked in a multidescription coding framework, cf., e.g., [12]. In addition to the layered approach for coherent and non-coherent communication in [3], receivers can access transmitted information at varying resolutions according to the SNR region in which they operate. Specifically, receivers operating in a lower SNR region are able to recover information from the basic layer, while those operating in a higher SNR region can recover information from the basic and refinement layers. Moreover, if CSI is available, any of the two classes can additionally recover coherent layer information.

IV. Multi-Layer-Based Detectors

In this section, we develop optimal and computationally-efficient suboptimal receivers. Before we do that, we write the $T \times N$ received signal as

$$\mathbf{Y} = (\alpha \mathbf{U V} + \beta \mathbf{U}_\perp) \, \mathbf{A H} + \frac{1}{\sqrt{\bar{\gamma}}} \, \mathbf{W}, \tag{5}$$

where $\bar{\gamma} = \gamma T/M$, $\mathbf{U}$ is drawn from the basic constellation, $\mathcal{C}_L$, and $\mathbf{V}$ and $\mathbf{A}$ are drawn from the refinement constellation, $\mathcal{C}_M$, and the coherent constellation, $\mathcal{C}_H$. The main priority of a multi-layer receiver is to retrieve the basic component $\mathbf{U}$, and if possible, it may retrieve either one or both of the refinement layer characterized by $\mathbf{V}$ and the coherent layer characterized by $\mathbf{A}$. This results in four classes of receivers, depending on the information recoverable by these receivers.

The following lemmas will be useful in subsequent analysis.

Lemma 1: Let $\mathbf{Y}$ be expressed as in (5). If $\mathbf{H}$ is unknown, then $\mathbf{Y}$ is statistically independent of $\mathbf{A}$.

Proof: The proof follows from the fact that unitarity of $\mathbf{A}$ implies that $\mathbf{AH}$ and $\mathbf{H}$ have the same distribution. ■

From this lemma, it can be seen that the performance of non-coherent receivers is not compromised by including $\mathbf{A}$ in the transmitted signal. Furthermore, it implies that the non-coherent receiver cannot benefit from knowing $\mathbf{A}$ nor can it use its codebook to simplify or improve the detection of non-coherent information. Next, we will obtain expression for the conditional distributions of the received signal.

Lemma 2: Let $\mathbf{Y}$ be expressed as in (5). Define $\mathbf{y} = \text{vec}\,(\mathbf{Y})$, and set

$$\mathbf{Q} = \alpha \mathbf{U V} + \beta \mathbf{U}_\perp, \tag{6}$$

then we have the following:

1) Conditioned on $\mathbf{U}$ and $\mathbf{V}$, the elements of $\mathbf{Y}$ are jointly Gaussian. Furthermore, $E(\mathbf{y}|\mathbf{U},\mathbf{V}) = 0$ and $\Gamma_{\mathbf{y}|\mathbf{U},\mathbf{V}} = E(\mathbf{y}\mathbf{y}^\dagger|\mathbf{U},\mathbf{V}) = \mathbf{I}\otimes\mathbf{Q}\mathbf{Q}^\dagger + 1/\bar{\gamma}\mathbf{I}$.
2) Conditioned on $\mathbf{U}$, $\mathbf{V}$, $\mathbf{A}$, and $\mathbf{H}$, the elements of $\mathbf{Y}$ are jointly Gaussian. Furthermore, $E(\mathbf{y}|\mathbf{U},\mathbf{V},\mathbf{A},\mathbf{H}) = \mathbf{QAH}$ and $\Gamma_{\mathbf{y}|\mathbf{U},\mathbf{V},\mathbf{A},\mathbf{H}} = 1/\bar{\gamma}\mathbf{I}$.
3) Conditioned on $\mathbf{U}$, the elements of $\mathbf{Y}$ are distributed as

$$p(\mathbf{Y}|\mathbf{U}) = \frac{\bar{\gamma}^{TN}}{|\mathcal{C}_M|\,\pi^{TN}(\bar{\gamma}+1)^{TN}}\times \sum_{\mathbf{Q}\in\{\mathbf{U}\}\times\mathcal{C}_M} \exp(-\operatorname{Tr}(\mathbf{Y}^\dagger(\bar{\gamma}\mathbf{I} - \frac{\bar{\gamma}^2}{\bar{\gamma}+1}\mathbf{Q}\mathbf{Q}^\dagger)\mathbf{Y})). \quad (7)$$

4) Conditioned on $\mathbf{U}$, $\mathbf{A}$, and $\mathbf{H}$, elements of $\mathbf{Y}$ are distributed as

$$p(\mathbf{Y}|\mathbf{U},\mathbf{A},\mathbf{H}) = \frac{\bar{\gamma}^{TN}}{|\mathcal{C}_M|\pi^{TN}}\times \sum_{\mathbf{Q}\in\{\mathbf{U}\}\times\mathcal{C}_M} \exp\left(-\bar{\gamma}\,\|\mathbf{Y}-\mathbf{QAH}\|^2\right). \quad (8)$$

Proof: Proof uses elementary conditioning arguments and is omitted for space limitations. ■

We now develop maximum likelihood (ML) and suboptimal detectors for each class of receivers.

A. Class A Receivers

This class comprises receivers that have no access to CSI and operate at lower SNRs. Receivers of this class can only detect the basic layer and their optimal ML detector decides on the codeword that solves the following optimization problem:

$$\max_{\mathbf{U}\in\mathcal{C}_L} \quad p(\mathbf{Y}|\mathbf{U}). \quad (9)$$

Drawing insight into $p(\mathbf{Y}|\mathbf{U})$, we note that the received signal expression in (5) can be expressed as

$$\mathbf{Y} = \alpha\mathbf{U}\mathbf{H}_2 + \beta\mathbf{U}_\perp\mathbf{H}_1 + \frac{1}{\sqrt{\bar{\gamma}}}\mathbf{W},$$

where elements of $\mathbf{H}_1 = \mathbf{AH}$ and $\mathbf{H}_2 = \mathbf{VAH}$ are both i.i.d zero mean Gaussian distributed. In addition, $\mathbf{H}_1$ and $\mathbf{H}_2$ are uncorrelated, however, not independent. This is because, despite being marginally Gaussian, $\mathbf{H}_1$ and $\mathbf{H}_2$ are not jointly Gaussian. To show that, we will use a contradiction argument. Suppose that $\mathbf{H}_1$ and $\mathbf{H}_2$ are jointly multivariate normal. A necessary condition that follows is that *every* deterministic affine transformation of $\begin{bmatrix}\mathbf{H}_1^\dagger & \mathbf{H}_2^\dagger\end{bmatrix}$ is also multivariate normal [13, Theorem 5.3.1]. Now, let us consider the matrix

$$\mathbf{B} = \begin{bmatrix}-\tilde{\mathbf{V}} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{H}_1\\ \mathbf{H}_2\end{bmatrix},$$

for some $\tilde{\mathbf{V}}\in\mathcal{C}_M$. Now, $\Pr(\mathbf{B}=0) = \Pr(\mathbf{V}-\tilde{\mathbf{V}}=0) = 1/|\mathcal{C}_M|$. This implies that the distribution of $\mathbf{B}$ has a delta function at the origin, and hence cannot be Gaussian.

Using (7) in (9), the ML detector will decide in favour of

$$\hat{\mathbf{U}}_{ML} = \arg\max_{\mathbf{U}\in\mathcal{C}_L} \sum_{\mathbf{V}\in\mathcal{C}_M} \exp\left(\frac{\bar{\gamma}^2}{\bar{\gamma}+1}\operatorname{Tr}\left(\mathbf{Y}^\dagger\mathbf{Q}\mathbf{Q}^\dagger\mathbf{Y}\right)\right), \quad (10)$$

where $\mathbf{Q}$ is defined in (6). The ML detector requires an exhaustive search over $|\mathcal{C}_L\times\mathcal{C}_M|$ points in order to find a point in a set of a notably-smaller size, namely $|\mathcal{C}_L|$. The discussion on a simplified detector in this case is deferred until the end of the next subsection, where simplified detectors for Classes A and B are proposed.

B. Class B Receivers

This class comprises receivers that do not have access to CSI but operate at higher SNRs. Such receivers will attempt to recover information in both the basic and refinement non-coherent layers. The ML detection is performed according to the rule

$$\left(\hat{\mathbf{U}}_{ML},\hat{\mathbf{V}}_{ML}\right) = \arg\max_{(\mathbf{U},\mathbf{V})\in\mathcal{C}_L\times\mathcal{C}_M} p(\mathbf{Y}|\mathbf{U},\mathbf{V}) = \arg\max_{\mathbf{Q}\in\mathcal{C}_L\times\mathcal{C}_M} p(\mathbf{Y}|\mathbf{Q}).$$

Using Lemma 2, the ML detector will decide in favour of

$$\hat{\mathbf{Q}}_{ML} = \arg\max_{\mathbf{Q}\in\mathcal{C}_L\times\mathcal{C}_M} \operatorname{Tr}\left(\mathbf{Y}^\dagger\mathbf{Q}\mathbf{Q}^\dagger\mathbf{Y}\right), \quad (11)$$

which is identical to the Generalized Likelihood Ratio Test (GLRT) detector in [14].

For this case, a simplified sequential detection scheme that exploits the underlying hierarchical structure was developed in [15]. In this scheme, the detection process is decomposed into two stages. Let $d:\mathbb{G}_{T,M}(\mathbb{C})\to\mathbb{R}$ be some decision metric and $q<|\mathcal{C}_L|$ be an integer. In the first stage, the detector identifies a candidate set of parent points, $\mathbf{U}\in\mathcal{C}_L$, with d greater than a constellation-dependent threshold τ, i.e.,

$$\mathcal{U} = \{\mathbf{U}\in\mathcal{C}_L|d(\mathbf{U})\geq\tau\}. \quad (12)$$

The threshold τ can be adjusted to ensure that $|\mathcal{U}| = q$. Herein, the decision metric is chosen to be the GLRT metric

$$d(\mathbf{U}) = \operatorname{Tr}\left(\mathbf{Y}^\dagger\mathbf{U}\mathbf{U}^\dagger\mathbf{Y}\right). \quad (13)$$

In the second stage, the detector examines only the children points of the parents in the candidate set, $\mathcal{U}$, and a decision is made in favor of

$$\hat{\mathbf{Q}} = \arg\max_{\mathbf{Q}\in\mathcal{U}\times\mathcal{C}_M} \operatorname{Tr}\left(\mathbf{Y}^\dagger\mathbf{Q}\mathbf{Q}^\dagger\mathbf{Y}\right). \quad (14)$$

Comparing this expression with (11) for Class B receivers (or (10) for Class A receivers), a reduction of $(|C_L|-q)\,|C_M|-|C_L|$ in ML computations is observed. Hence, ensuring that q is small implies significant computational savings.

C. Class C Receivers

This class encompasses receivers with perfect CSI, operating at lower SNRs. This class will be able to coherently recover information in the basic and coherent layers. Hence, the ML rule in this case can expressed as

$$\left(\hat{\mathbf{U}}_{ML},\hat{\mathbf{A}}_{ML}\right) = \max_{(\mathbf{U},\mathbf{A})\in\mathcal{C}_L\times\mathcal{C}_H} p(\mathbf{Y}|\mathbf{U},\mathbf{A},\mathbf{H}).$$

Using (8), it follows that the ML detection rule reduces to

$$\left(\hat{\mathbf{U}}_{ML},\hat{\mathbf{A}}_{ML}\right) = \max_{(\mathbf{U},\mathbf{A})}\sum_{\mathbf{V}\in\mathcal{C}_M}\exp(-\bar{\gamma}\|\mathbf{Y}-\mathbf{QAH}\|^2).$$

where $\mathbf{Q}$ is given in (6). Similar to the Class A ML detector, the search space size is increased by a factor of $|\mathcal{C}_M|$. To overcome this limitation, we will propose two simplified detection schemes: non-sequential and sequential ones.

For the non-sequential scheme, let $\mathbf{Z} = \mathbf{U}_\perp^\dagger \mathbf{Y}$. Then,

$$\mathbf{Z} = \beta \mathbf{A}\mathbf{H} + \frac{1}{\sqrt{\bar{\gamma}}}\mathbf{W}. \tag{15}$$

Rather than maximizing $p(\mathbf{Y}|\mathbf{U}, \mathbf{A}, \mathbf{H})$, the simplified detector performs the ML test on $\mathbf{Z}$. Conditioned on $\mathbf{U}$, $\mathbf{A}$ and $\mathbf{H}$, the elements of $\mathbf{Z}$ are jointly Gaussian. Hence, the simplified detection rule is

$$(\hat{\mathbf{U}}, \hat{\mathbf{A}}) = \min_{(\mathbf{U},\mathbf{A})\in\mathcal{C}_L\times\mathcal{C}_H} \left\| \mathbf{U}_\perp^\dagger \mathbf{Y} - \beta\mathbf{A}\mathbf{H} \right\|. \tag{16}$$

It is crucial to point out that $\mathbf{Z}$ contains only side information about $\mathbf{A}$. In fact, one can show that the proposed detector is strictly sub-optimal by showing that the mutual information $I(\mathbf{A};\mathbf{Y}) > I(\mathbf{A};\mathbf{Z})$.

In terms of computational cost, the simplified detector requires search over $|\mathbf{C}_L||\mathbf{C}_H|$ as opposed to $|\mathbf{C}_L||\mathbf{C}_M||\mathbf{C}_H|$. However, even for moderate-size constellations, the number of computations involved can be massive and may result in unnecessary processing overhead.

For the sequential scheme, we reuse the discussion on the simplified decoder in the previous section. Specifically, a candidate parent set $\mathcal{U}$ of reduced size q is first generated according to the distance metric in (13). Next, the likelihood search in (16) is performed over $\mathcal{U} \times \mathcal{C}_H$ in order to find the decision pair $(\hat{\mathbf{U}}, \hat{\mathbf{A}})$ that minimizes the objective in (16).

D. Class D Receivers

Receivers of this class operate coherently and at the higher SNRs. Therefore, these receivers can detect all transmitted layers. The ML estimate is in this case is given by

$$\begin{aligned}(\hat{\mathbf{Q}}_{ML}, \hat{\mathbf{A}}_{ML}) &= \max_{(\mathbf{Q},\mathbf{A})\in\mathcal{C}_L\times\mathcal{C}_M\times\mathcal{C}_H} p(\mathbf{Y}|\mathbf{Q},\mathbf{A},\mathbf{H}) \\ &= \min_{(\mathbf{Q},\mathbf{A})\in\mathcal{C}_L\times\mathcal{C}_M\times\mathcal{C}_H} \|\mathbf{Y} - \mathbf{Q}\mathbf{A}\mathbf{H}\|. \end{aligned} \tag{17}$$

A three-step method for simplified detection can be readily deduced in the following manner. The first two steps parallel the operation of the sequential non-coherent detector for Class B receivers, where the basic and refinement information are non-coherently estimated. Next, the coherent component is estimated coherently based on the decision taken on the non-coherent component and knowledge of $\mathbf{H}$. That is, the decision in the last step is made according to the following rule:

$$\hat{\mathbf{A}} = \min_{\mathbf{A}\in\mathcal{C}_H} \|\mathbf{Y} - \hat{\mathbf{Q}}\mathbf{A}\mathbf{H}\|. \tag{18}$$

V. Performance Evaluation

In this section we analyze the performance of the proposed multi-step detectors presented in the previous section. Towards this end, we plot the symbol error rates (SER) of the simplified detectors against those of the optimal detection schemes. In all cases, we assume $T = 4$, and $M = N_i = 2$. The overall constellation size is $2^{11} = 2048$, which is partitioned into

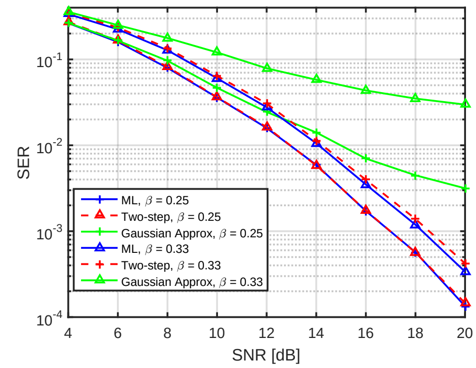

Fig. 2. SER of basic layer for ML, two-step, and Gaussian detectors.

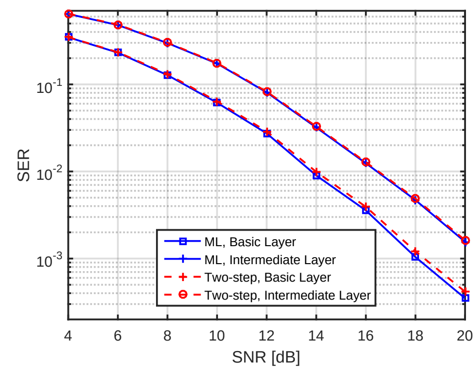

Fig. 3. SER of basic and refinement layers for ML and two-step detectors.

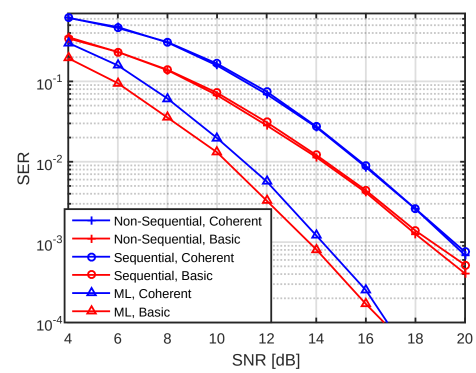

Fig. 4. SER of basic and coherent layers for ML, non-sequential, and sequential two-step detectors.

$16 \times 8 \times 8$ for the basic, refinement and coherent layers, respectively. The respective rates are $1, 0.75, 0.75$ bits per channel use (bpcu). Throughout, the size of the candidate parent set for the simplified detectors, $\mathcal{U}$, is $q = 2$, resulting in a computational saving of about 87%. The coherent code used throughout is the standard 2×2 Alamouti one,

$$\mathbf{A} = \frac{1}{\sqrt{2}} \begin{bmatrix} x_1 & x_2 \\ -x_2^* & x_1^* \end{bmatrix},$$

where x_1 and x_2 are drawn from a 4-QAM constellation.

1) Class A Receivers: For these receivers, we consider two cases: $\beta = 0.25$ and $\beta = 0.33$. In Fig. 2, we plot the SER for the optimal and the simplified two-step detector. The SER of the ML detector assuming that the distribution $p(\mathbf{Y}|\mathbf{U})$

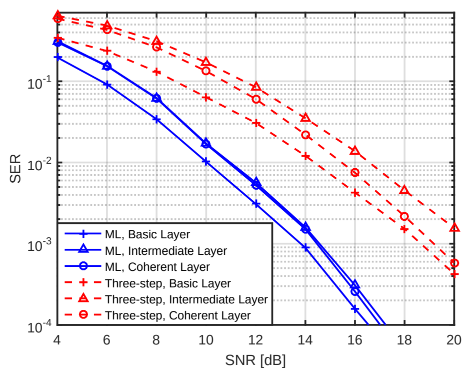

Fig. 5. SER of all three layers for ML and simplified three-step detectors.

is Gaussian is also shown. For a zero-mean refinement layer and assuming that the conditional distribution is Gaussian, the expression for the ML detection rule is

$$\max_{\mathbf{U}\in\mathcal{C}_L} \operatorname{Tr}(\mathbf{Y}^\dagger\mathbf{U}\mathbf{U}^\dagger\mathbf{Y}).$$

It can be seen from this figure that the sequential scheme exhibits almost identical performance to that of the optimal detector. Additionally, we observe heavy performance losses for the ML detector when the distribution in (7) is erroneously assumed to be Gaussian. Furthermore, for this detector, the gap from optimality grows substantially with the increase of β. This observation can be intuitively explained: Assuming the distribution is Gaussian, the ML decoder attempts to find the transmitted child point by examining parent points. When β is increased, children move further away from their associated parents, thus leading to this deterioration.

2) Class B Receivers: In Fig. 3, the SER for the basic, as well as the refinement layer, is shown. Again, in this case, we observe a near-optimal performance of the simplified approach, despite the considerable reduction in ML computations.

3) Class C Receivers: The error rate probabilities are investigated for the sequential and non-sequential detectors described in Section IV-C. The SER of the simplified detectors is compared against that of the optimal ML detector in Fig. (4). From this figure, we draw two main observations. First, the non-sequential detector provides no noticeable advantage over the sequential counterpart. Second, the ML detector outperforms the simplified detectors by about 4 dB for the basic layer and 5 dB for the coherent layer at an SER of 10^{-3}. We conclude that ignoring side-information, due to the projection $\mathbf{Z} = \mathbf{U}_\perp^\dagger\mathbf{Y}$, is the cause for this deterioration in performance.

4) Class D Receivers: We compare the SER in all three layers for the three-step detector with that of the one-step optimal detector in Fig. (5). Again, we observe a tangible loss in performance for the simplified detector relative to the optimal one, e.g., 5, 5 and 6 dB for the basic, refinement and coherent layers, respectively, at an SER of 10^{-3}. This deterioration follows from the non-coherent detection of the basic and refinement layers. In other words, in the detection of those layers, the three-step detector does not make use of the available CSI.

VI. Conclusion

In this paper, we developed a multi-layer approach comprised of a two-layer non-coherent component and a one-layer coherent one. These components are mutually utilized to convey basic LR and refinement HR information. The receivers can be categorized into four classes depending on CSI availability and operational SNR. Each of these classes will have access to a particular subset of the transmitted layers. For each of these classes, the ML detection criterion is derived and, simplified detectors are developed to overcome the computational limitations of ML detectors. Numerical simulations were used to show that the simplified detectors enable comparable, and in some cases identical, performance to the ML counterparts.

References

[1] J. M. Shea, K. Sistla, and B. A. Davis, "Multicasting in the forward link of cdma cellular systems," in *MILCOM 2002. Proceedings*, vol. 2, Oct 2002, pp. 923–927 vol.2.

[2] J. Liu and A. Annamalai, "Multi-resolution signaling for multimedia multicasting," in *IEEE 60th Vehicular Technology Conference, 2004. VTC2004-Fall. 2004*, vol. 2, Sept 2004, pp. 1088–1092 Vol. 2.

[3] K. G. Seddik, R. H. Gohary, M. T. Hussien, M. Shaqfeh, H. Alnuweiri, and H. Yanikomeroglu, "Multi-resolution multicasting over the grassmann and stiefel manifolds," *IEEE Transactions on Wireless Communications*, vol. 16, no. 8, pp. 5296–5310, Aug 2017.

[4] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM Journal on Matrix Analysis and Applications*, vol. 20, no. 2, pp. 303–353, 1998.

[5] J. H. Conway, R. H. Hardin, and N. J. Sloane, "Packing lines, planes, etc.: Packings in grassmannian spaces," *Experimental Mathematics*, vol. 5, no. 2, pp. 139–159, 1996.

[6] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 543–564, 2000.

[7] L. Zheng and D. N. Tse, "Communication on the grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 359–383, 2002.

[8] R. H. Gohary and T. N. Davidson, "Noncoherent mimo communication: Grassmannian constellations and efficient detection," *Information Theory, IEEE Transactions on*, vol. 55, no. 3, pp. 1176–1205, 2009.

[9] D. Argawal, T. Richardson, and R. Urbanke, "Multiple-antenna signal constellations for fading channel," *IEEE Transactions on Informatiion Theory*, vol. 47, pp. 2618–2626, 2001.

[10] K. M. Attiah, K. Seddik, R. H. Gohary, and H. Yanikomeroglu, "A systematic design approach for non-coherent grassmannian constellations," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 2948–2952.

[11] I. S. Dhillon, J. R. Heath, T. Strohmer, and J. A. Tropp, "Constructing packings in grassmannian manifolds via alternating projection," *Experimental Mathematics*, vol. 17, no. 1, pp. 9–35, 2008.

[12] V. K. Goyal, "Multiple description coding: compression meets the network," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 74–93, Sep 2001.

[13] A. Gut, *An Intermediate Course in Probability*, 2nd ed. Springer Publishing Company, Incorporated, 2009.

[14] M. Brehler and M. K. Varanasi, "Asymptotic error probability analysis of quadratic receivers in rayleigh-fading channels with applications to a unified analysis of coherent and noncoherent space-time receivers," *IEEE Transactions on Information Theory*, vol. 47, no. 6, pp. 2383–2399, 2001.

[15] K. M. Attiah, K. Seddik, R. H. Gohary, and H. Yanikomeroglu, "Non-coherent multi-layer constellations for unequal error protection," in *2017 IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–6.