MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

High-Order Super-Block GLRT for Non-Coherent Grassmann Codes in MIMO-OFDM Systems

Koike-Akino, T.; Orlik, P.

TR2010-120 December 2010

Abstract

We investigate non-coherent multi-input multi-output (MIMO) signal processing which requires no channel state information (CSI) at either the transmitter or the receiver. With non-coherent codes on Grassmann manifold, a receiver employing generalized likelihood ratio test (GLRT) algorithm offers the maximum-likelihood performance even without CSI. However, the conventional GLRT suffers from a severe performance degradation when the channel changes fast within a coding block duration.We propose an improved GLRT algorithm referred to as high-order super-block techniques. The super-block scheme makes effective use of correlated channels for adjacent blocks in slow fading, whereas the high-order scheme can overcome the channel fluctuation during a block in fast fading. We demonstrate that the proposed scheme significantly improves performance for MIMO-OFDM with non-coherent Grassmann space-frequency block codes (SFBC).

IEEE Global Telecommunications Conference (GLOBECOM)

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE Globecom 2010 proceedings.

High–Order Super–Block GLRT for Non–Coherent Grassmann Codes in MIMO–OFDM Systems

Toshiaki Koike-Akino*† and Philip Orlik†

* School of Engineering and Applied Sciences, Harvard University, 33 Oxford Street, Cambridge, MA 02138, U.S.A.

† Mitsubishi Electric Research Labs (MERL), 201 Broadway, Cambridge, MA 02139, U.S.A.

Email: koike@seas.harvard.edu, {koike, porlik}@merl.com

***Abstract*—We investigate non–coherent multi–input multi–output (MIMO) signal processing which requires no channel state information (CSI) at either the transmitter or the receiver. With non–coherent codes on Grassmann manifold, a receiver employing generalized likelihood ratio test (GLRT) algorithm offers the maximum–likelihood performance even without CSI. However, the conventional GLRT suffers from a severe performance degradation when the channel changes fast within a coding block duration. We propose an improved GLRT algorithm referred to as high–order super–block techniques. The *super–block* scheme makes effective use of correlated channels for adjacent blocks in slow fading, whereas the *high–order* scheme can overcome the channel fluctuation during a block in fast fading. We demonstrate that the proposed scheme significantly improves performance for MIMO–OFDM with non–coherent Grassmann space–frequency block codes (SFBC).**

I. Introduction

A large number of studies in the past dozen years have proven that the use of multiple antennas at both transmitter and receiver, the technique of which is known as multi–input multi–output (MIMO) systems, can dramatically increase data throughput. In rich–scattering channel environments, the link capacity increases linearly with $\min(M, N)$ [1, 2], where M and N denote the number of transmitting antennas and that of receiving antennas, respectively. However, an accurate channel state information (CSI) is required for achieving such capacity gains.

Without CSI, we require *non–coherent* communications. Some information–theoretical works [3, 4] have verified that the non–coherent channel capacity becomes a function of $M'(1 - M'/L)$ in high signal–to–noise ratio (SNR) regimes, where $M' = \min(M, N, \lfloor L/2 \rfloor)$ and L denotes the length of non–coherent codes with $\lfloor x \rfloor$ being the floor function. Those theoretical studies have motivated several works on signal design of non–coherent codes, which include unitary space–time constellations [5–7], exponential mapping Grassmann codes [8, 9], non–parametric Grassmann codes [10, 11], and differential space–time modulations [12, 13].

Marzetta and Hochwald showed in [3] that unitary space–time codes asymptotically achieve the non–coherent channel capacity for high SNRs. For such codes, the optimal performance of maximum–likelihood decoding can be offered by a generalized likelihood ratio test (GLRT) [14] even without CSI. The key idea behind the GLRT receiver lies in the fact that it employs implicit channel estimation for each codeword of the non–coherent codes at the time of decoding. However, the performance of the conventional GLRT receiver seriously degrades when the channel coherence length (in time or frequency domain) is shorter than the length of non–coherent codes L. To prevent such a performance degradation, the code length should be reasonably short in practice. Shorter space–time codes in turn decrease the capacity gains of $M'(1 - M'/L)$. In this paper, we propose an improved GLRT receiver in order to deal with the above–mentioned tradeoff between performance and the code length for any arbitrary channel coherence length.

The proposed GLRT incorporates two novel ideas; i) *super–block* scheme, which can improve performance for short code lengths in slow fading channels by considering consecutive coded blocks as one extended coding block, and ii) *high–order* scheme, which can overcome the performance degradation in rapidly varying fading channels over a code block duration by adopting high–order least–squares (LS) regressions. For the super–block GLRT, we further propose an improved decoding algorithm referred to as sequential decision which is based on the Viterbi algorithm to efficiently decode consecutive space–time block codes. In this paper, we demonstrate that our proposed scheme offers a significant improvement in error rate performance for non–coherent MIMO–OFDM systems.

Notations: Throughout the paper, we describe matrices and vectors by bold–face italic letters in capital cases and small cases, respectively. Let $\boldsymbol{X} \in \mathbb{C}^{m \times n}$ be a complex–valued $(m \times n)$–dimensional matrix, where $\mathbb{C}$ denotes the complex field. The notations $\boldsymbol{X}^*$, $\boldsymbol{X}^{\mathrm{T}}$, $\boldsymbol{X}^{\dagger}$, $\boldsymbol{X}^{-1}$, $\mathrm{tr}[\boldsymbol{X}]$ and $\|\boldsymbol{X}\|$ represent the complex conjugate, the transpose, the Hermite transpose, the inverse, the trace and the Frobenius norm of $\boldsymbol{X}$, respectively. The operator $\mathrm{vec}[\cdot]$ denotes the vector–operation which stacks all columns of a matrix into a single column vector in a left–to–right fashion, and the operator $\otimes$ stands for the Kronecker product of two matrices. The set of real numbers is denoted by $\mathbb{R}$, and $\boldsymbol{I}_m \in \mathbb{R}^{m \times m}$ denotes an m–dimensional identity matrix.

II. Non–Coherent MIMO Signal Processing

A. Non–Coherent MIMO–OFDM Systems

We consider $M \times N$ MIMO–OFDM systems in which M transmitting antennas and N receiving antennas are used.

We focus on non–coherent communications where both the transmitter and the receiver do not have any CSI. The use of non–coherent codes [5–13] enables us to communicate efficiently without pilot or training sequences for channel estimations.

Let $\boldsymbol{x}_n \in \mathbb{C}^{M\times 1}$ be the signal vector transmitted from M antennas at the n–th subcarrier. We assume L–symbol block of one codeword. Each block consists of $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_L] \in \mathbb{X} \subset \mathbb{C}^{M\times L}$, where $\mathbb{X} = \{\boldsymbol{\mathcal{X}}_1, \boldsymbol{\mathcal{X}}_2, \ldots, \boldsymbol{\mathcal{X}}_Q\}$ denotes a non–coherent space–frequency codebook with Q distinct codewords. The mean energy of the codeword per transmit antenna is normalized, namely $(1/QM)\sum_{q=1}^{Q}\|\boldsymbol{\mathcal{X}}_q\|^2 = 1$.

Over MIMO channels, the received signal is written as

$$\boldsymbol{y}_n = \boldsymbol{H}_n \boldsymbol{x}_n + \boldsymbol{w}_n, \tag{1}$$

where $\boldsymbol{y}_n \in \mathbb{C}^{N\times 1}$, $\boldsymbol{H}_n \in \mathbb{C}^{N\times M}$ and $\boldsymbol{w}_n \in \mathbb{C}^{N\times 1}$ denote the received signals vector, the (frequency–domain) MIMO channel matrix and the additive noise, respectively, at the n–th subcarrier. In order to describe the conventional GLRT receiver [14], we first assume that the MIMO channel matrix remains constant such that $\boldsymbol{H}_n = \boldsymbol{H}$ during a single block for $n = 1, 2, \ldots, L$ (this assumption will be relaxed later when we introduce a high–order scheme to deal with channel fluctuations over a block length). This assumption of block fading channels simplifies the expression of the received signal into a matrix form as follows:

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{W}, \tag{2}$$

where $\boldsymbol{Y}$ and $\boldsymbol{W}$ denote the received signals and the additive noise signals over the code block:

$$\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_L] \in \mathbb{C}^{N\times L}, \tag{3}$$
$$\boldsymbol{W} = [\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_L] \in \mathbb{C}^{N\times L}. \tag{4}$$

We suppose that the noise is white Gaussian random variables: $\mathbb{E}[\mathrm{vec}[\boldsymbol{W}]\mathrm{vec}[\boldsymbol{W}]^\dagger] = \sigma^2 \boldsymbol{I}_{NL}$. Note that it is straightforward to consider a colored noise as discussed in [11].

B. Generalized Likelihood Ratio Test (GLRT) Receiver

The conditional probability of $\boldsymbol{Y}$ given $\boldsymbol{X}$ and $\boldsymbol{H}$ is known as the likelihood which is expressed as

$$\Pr(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{H}) = \frac{1}{(\pi\sigma^2)^{NL}} \exp\left(-\frac{1}{\sigma^2}\|\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{X}\|^2\right).$$

Without CSI, the GLRT receiver [14] searches for the best estimate $\hat{\boldsymbol{X}}$ from the codebook $\mathbb{X}$ in favor of maximizing the likelihood, or equivalently minimizing the squared distance metric as follows:

$$\hat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}\in\mathbb{X}} \inf_{\boldsymbol{H}} \|\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{X}\|^2. \tag{5}$$

Note that since $\boldsymbol{H}$ is not known at the receiver, the GLRT uses the best channel matrix over all the possible realizations for each codewords.

Since we have

$$\frac{\partial}{\partial \boldsymbol{H}^*}\|\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{X}\|^2 = -\left(\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{X}\right)\boldsymbol{X}^\dagger, \tag{6}$$

the channel candidate $\hat{\boldsymbol{H}} = \boldsymbol{Y}\boldsymbol{X}^\dagger(\boldsymbol{X}\boldsymbol{X}^\dagger)^{-1}$ can maximize the likelihood, where we assume $\boldsymbol{X}\boldsymbol{X}^\dagger$ is invertible. This is equivalent to the well–known lest–squares (LS) channel estimation given a codeword candidate $\boldsymbol{X}$. Substituting $\hat{\boldsymbol{H}}$ for $\boldsymbol{H}$ in (5) yields

$$\hat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}\in\mathbb{X}} \Big\|\boldsymbol{Y}\underbrace{\left(\boldsymbol{I}_L - \boldsymbol{X}^\dagger(\boldsymbol{X}\boldsymbol{X}^\dagger)^{-1}\boldsymbol{X}\right)}_{\boldsymbol{P}}\Big\|^2. \tag{7}$$

If every codeword is orthonormal such that $\boldsymbol{\mathcal{X}}_q\boldsymbol{\mathcal{X}}_q^\dagger = \boldsymbol{I}_M$ for any $q = 1, 2, \ldots, Q$, the GLRT metric can be further simplified to $\max\|\boldsymbol{Y}\boldsymbol{X}^\dagger\|^2$.

Here, a matrix $\boldsymbol{P} \in \mathbb{P} \subset \mathbb{C}^{L\times L}$ denotes an idempotent projector onto the orthogonal complement of a codeword $\boldsymbol{X}$; i.e., $\boldsymbol{X}\boldsymbol{P} = \boldsymbol{0}$ and $\boldsymbol{P}\boldsymbol{P} = \boldsymbol{P}$. The set $\mathbb{P} = \{\boldsymbol{\mathcal{P}}_1, \boldsymbol{\mathcal{P}}_2, \ldots, \boldsymbol{\mathcal{P}}_Q\}$ is a projector bank, whose q–th member is defined as $\boldsymbol{\mathcal{P}}_q = \boldsymbol{I}_L - \boldsymbol{\mathcal{X}}_q^\dagger(\boldsymbol{\mathcal{X}}_q\boldsymbol{\mathcal{X}}_q^\dagger)^{-1}\boldsymbol{\mathcal{X}}_q$, for the codebook $\mathbb{X}$. It should be noted that the minimum size of the possible projector matrix $\boldsymbol{P}$ such that $\boldsymbol{X}\boldsymbol{P} = \boldsymbol{0}$ can be $L\times(L-M)$ because the orthogonal complement of $\boldsymbol{X}$ is of size $L\times(L-M)$. More importantly, given a codebook $\mathbb{X}$, the projector bank $\mathbb{P}$ is determined in advance.

C. Non–Coherent Grassmann Codes

A number of non–coherent codes have been reported, e.g., unitary space–time codes [5–7, 13], Grassmann codes with exponential mapping [8, 9], Grassmann packing codes with numerical optimization [10, 11], and differential modulations [12]. Here, we describe a non–coherent Grassmann code based on the exponential mapping design presented in [8, 9]. The exponential mapping technique enables us to design a good non–coherent codes from coherent codes in a straightforward manner while the full–rate and full–diversity gains hold.

Grassmann codes parameterized by the exponential mapping technique are written as

$$\boldsymbol{X} = \begin{bmatrix}\boldsymbol{I}_M & \boldsymbol{0}_{M\times(L-M)}\end{bmatrix}\exp\left(\begin{bmatrix}\boldsymbol{0}_M & \boldsymbol{B}\\ -\boldsymbol{B}^\dagger & \boldsymbol{0}_{L-M}\end{bmatrix}\right),$$

The matrix $\boldsymbol{B} \in \mathbb{C}^{M\times(L-M)}$ denotes a full–rate full–diversity coherent space–time block code. Let $\boldsymbol{B} = \boldsymbol{U}\boldsymbol{A}\boldsymbol{V}^\dagger$ be the singular value decomposition (SVD) where $\boldsymbol{U} \in \mathbb{C}^{M\times M}$ and $\boldsymbol{V} \in \mathbb{C}^{(L-M)\times(L-M)}$ are unitary matrices and $\boldsymbol{A} \in \mathbb{C}^{M\times(L-M)}$ is a diagonal matrix. The cosine–sine decomposition yields

$$\boldsymbol{X} = \begin{bmatrix}\boldsymbol{U}\cos(\alpha\boldsymbol{A})\boldsymbol{U}^\dagger & \boldsymbol{U}\sin(\alpha\boldsymbol{A})\boldsymbol{V}^\dagger\end{bmatrix}, \tag{8}$$

where α is a parameter which controls the codeword distance. Such a codeword always satisfies the orthonormal condition, $\boldsymbol{X}\boldsymbol{X}^\dagger = \boldsymbol{I}_M$, for any arbitrary α and $\boldsymbol{B}$.

For $M = 2$ and $L = 4$, one choice of the coherent coding matrix $\boldsymbol{B}$ suggested in [8, 9] is

$$\boldsymbol{B} = \frac{1}{\sqrt{2}}\begin{bmatrix} s_1 + \phi s_2 & \vartheta(s_3 + \phi s_4)\\ \vartheta(s_3 - \phi s_4) & s_1 - \phi s_2\end{bmatrix}, \tag{9}$$

where $\vartheta^2 = \phi = \exp(\jmath\pi/4)$. The optimal α, in this case, is approximately 0.566. Each s_i is drawn from 4QAM constellations for a spectral efficiency of 2 bps per channel use. In [8, 9], it has been shown that this Grassmann code offers the maximum degrees of freedom (multiplexing gain) for non–coherent communications. However, it is not obvious whether such a coherent coding matrix $\boldsymbol{B}$ with the parameters ϑ and ϕ provides the best sphere packing performance over the Grassmann manifold. We will later present better Grassmann codes with parameters optimized by a gradient method.

III. High–Order Super–Block GLRT

A. Super–Block GLRT

The length of the non–coherent codes L should be shorter than the coherence bandwidth in principle. However, shorter space–frequency codes have a poor performance with the conventional GLRT receiver. It is because the accuracy of the LS channel regressions decreases linearly with the code length L. Even for highly selective fading channels (in frequency–domain for SFBC), the channel matrix has a high correlation for adjacent blocks in general. It suggests that we can enjoy a performance gain through the use of channel correlations when we increase the effective block length by coupling multiple blocks at receivers.

We propose super–block GLRT receiver, which jointly estimates K adjacent blocks. Let $\boldsymbol{X}_k$ and $\boldsymbol{Y}_k$ be the transmission block and the received block at the k–th block for $k \in \{1, 2, \ldots, K\}$. When we write a super–block received signal and a super–block transmitted signal as

$$\boldsymbol{Y}' = [\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_K] \in \mathbb{C}^{N\times LK}, \tag{10}$$
$$\boldsymbol{X}' = [\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_K] \in \mathbb{C}^{M\times LK}, \tag{11}$$

we can use the GLRT receiver if the channel remains constant over these K blocks. Here, the signal $\boldsymbol{X}'$ is a new virtual codebook generated from the original codebook $\boldsymbol{X}_k \in \mathbb{X}$. The corresponding projector matrix $\boldsymbol{P}'$, as in (7), can be computed in advance such that $\boldsymbol{X}'\boldsymbol{P}' = \boldsymbol{0}$. Note that the computational complexity increases exponentially with the number of blocks, K, because the cardinality of a super–block codebook becomes $Q' = Q^K$. If we use orthonormal codes, the GLRT metric reduces to $\max \| \sum_{k=1}^{K} \boldsymbol{Y}_k \boldsymbol{X}_k^\dagger \|^2$.

B. Sequential Decision for Super–Block GLRT

Since the super–block GLRT treats multiple blocks at the same time, some different decision criteria arise as follows. Let $\mu_j = \left\| \left[\boldsymbol{Y}_{j+1}, \boldsymbol{Y}_{j+2}, \ldots, \boldsymbol{Y}_{j+K}\right] \boldsymbol{P}' \right\|^2$ be the metric of the super–block GLRT for K consecutive blocks from $\boldsymbol{X}_{j+1}$ to $\boldsymbol{X}_{j+K}$.

- *One–time decision*: use only the metric $\mu_{\lfloor k/K \rfloor -1}$.
- *Selective decision*: select the best metric out of the adjacent metrics from μ_{k-K+1} to μ_{k+K-1}.
- *Combined decision*: use the combined metric which is summed up all the metrics from μ_{k-K+1} to μ_{k+K-1}.
- *Sequential decision*: select the best metrics all over the blocks by using Viterbi algorithm. The previous $K-1$ blocks are interpreted as trellis states. Along the trellis–state diagram, optimal decision can be obtained.

Obviously, sequential decision has the highest complexity while achieving the best performance.

C. High–Order Super–Block GLRT

The GLRT receiver in principle requires an assumption that the channel remains static during a super block (or, consecutive LK symbols). Hence, a channel fluctuation during a super block may incur a severe performance degradation. Here, we propose an improved GLRT which uses high–order LS channel estimation [15] in order to overcome the channel variation during the block. Let us use the D–th order polynomial curves to fit the channel fluctuation for high–order LS regressions. The channel matrix at the n–th subcarrier is then modeled as follows:

$$\boldsymbol{H}_n = \sum_{d=0}^{D} \boldsymbol{H}^{[d]} n^d = \boldsymbol{\mathcal{H}} \boldsymbol{D}_n, \tag{12}$$

where

$$\boldsymbol{\mathcal{H}} = [\boldsymbol{H}^{[0]}, \boldsymbol{H}^{[1]}, \ldots, \boldsymbol{H}^{[D]}] \in \mathbb{C}^{N\times M(D+1)}, \tag{13}$$
$$\boldsymbol{D}_n = [n^0 \boldsymbol{I}_M, n^1 \boldsymbol{I}_M, \ldots, n^D \boldsymbol{I}_M]^{\mathrm{T}} \in \mathbb{R}^{M(D+1)\times M}. \tag{14}$$

The matrix $\boldsymbol{H}^{[d]}$ denotes the channel matrix at the d–th order term of the polynomial for $d \in \{0, 1, \ldots, D\}$. This model enables us to adopt the GLRT receiver even when $\boldsymbol{H}_n$ is changing in the frequency domain because the expanded channel matrix $\boldsymbol{\mathcal{H}}$ remains static.

The received signal can be rewritten as

$$\boldsymbol{Y}' = \boldsymbol{\mathcal{H}} \underbrace{\boldsymbol{\mathcal{D}} \boldsymbol{\Gamma}}_{\boldsymbol{X}'} + \boldsymbol{W}, \tag{15}$$

where $\boldsymbol{\mathcal{D}}$ is the (deterministic) order expansion matrix and $\boldsymbol{\Gamma}$ is the diagonally aligned version of the transmitted signal matrix $\boldsymbol{X}$, each of which is defined as

$$\boldsymbol{\mathcal{D}} = [\boldsymbol{D}_1, \boldsymbol{D}_2, \ldots, \boldsymbol{D}_{LK}] \in \mathbb{R}^{M(D+1)\times MLK}, \tag{16}$$

$$\boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{x}_1 & & & \\ & \boldsymbol{x}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{x}_{LK} \end{bmatrix} \in \mathbb{C}^{MLK\times LK}. \tag{17}$$

By considering $\boldsymbol{X}' = \boldsymbol{\mathcal{D}}\boldsymbol{\Gamma} \in \mathbb{C}^{M(D+1)\times LK}$ as a new virtual codeword, the associated projector matrix becomes

$$\boldsymbol{P}' = \boldsymbol{I}_{LK} - \boldsymbol{\Gamma}^\dagger \boldsymbol{\mathcal{D}}^\dagger (\boldsymbol{\mathcal{D}}\boldsymbol{\Gamma}\boldsymbol{\Gamma}^\dagger\boldsymbol{\mathcal{D}}^\dagger)^{-1} \boldsymbol{\mathcal{D}}\boldsymbol{\Gamma} \in \mathbb{C}^{LK\times LK}, \tag{18}$$

which can be computed in advance for any D and for all codewords.

It should be noted that the computational complexity of the high–oder GLRT is independent of the order D because the size of predetermined projector matrix $\boldsymbol{P}'$ does not increase. There is a constraint on the maximum available order D; more specifically, $M(D+1) < LK$ must be fulfilled because $M(D+1) > LK$ results in rank deficiency of the term $\boldsymbol{\mathcal{D}}\boldsymbol{\Gamma}\boldsymbol{\Gamma}^\dagger\boldsymbol{\mathcal{D}}^\dagger$. This drawback can be dealt with by increasing the super–block length K. Obviously, higher–order modelling is advantageous only if the channel response frequently changes during a super block.

IV. Codebook Optimization of Non–Coherent Grassmann Codes

In this section, we optimize non–coherent codes numerically in a similar way of [11]. The optimization of non–coherent codes is done by sphere packing on the Grassmann manifold. For numerical Grassmann packing, we adopt the gradient method to minimize the pairwise error probability between two codewords in high SNR regimes.

A. Design Criterion: Pairwise Error Probability

As shown in [11], we can write the pairwise error probability (between the correct codeword $\boldsymbol{\mathcal{X}}_i$ and the wrong codeword $\boldsymbol{\mathcal{X}}_j$) given a channel matrix $\boldsymbol{H}$ as

$$\begin{aligned}
&\Pr(\boldsymbol{\mathcal{X}}_i \to \boldsymbol{\mathcal{X}}_j \mid \boldsymbol{H}) = \Pr(\|\boldsymbol{Y}\boldsymbol{\mathcal{P}}_j\|^2 < \|\boldsymbol{Y}\boldsymbol{\mathcal{P}}_i\|^2 \mid \boldsymbol{H}) \\
&\simeq \Pr(\|\boldsymbol{H}\boldsymbol{\mathcal{X}}_i\boldsymbol{\mathcal{P}}_j\|^2 + 2\Re[\mathrm{tr}[\boldsymbol{H}\boldsymbol{\mathcal{X}}_i\boldsymbol{\mathcal{P}}_j\boldsymbol{W}^\dagger]] < 0 \mid \boldsymbol{H}) \\
&= \frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{\|\boldsymbol{H}\boldsymbol{\mathcal{X}}_i\boldsymbol{\mathcal{P}}_j\|^2}{4\sigma^2}}\right) \le \frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{\|\boldsymbol{H}\|^2\lambda_{\min}(\boldsymbol{\mathcal{X}}_i\boldsymbol{\mathcal{P}}_j\boldsymbol{\mathcal{X}}_i^\dagger)}{4\sigma^2}}\right),
\end{aligned}$$

in the high SNR regimes, where $\mathrm{erfc}(\cdot)$ is the complementary error function, and $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue of a matrix.

Our goal is to design a codebook $\mathbb{X}$ which maximizes $\lambda_{\min}(\boldsymbol{\mathcal{X}}_i\boldsymbol{\mathcal{P}}_j\boldsymbol{\mathcal{X}}_i^\dagger)$ for any possible pair $i \neq j$. In [11], a semi–definite programming (SDP) relaxation method is used for optimization with an energy constraint. The codebook obtained by SDP is further refined by a linear programming method. In this paper, we optimize the codebook for the high–order super–block GLRT by the gradient method as this is a lower–complexity approach.

B. Gradient Method

For given $\boldsymbol{\Omega}_{i,j} = \boldsymbol{\mathcal{X}}_i\boldsymbol{\mathcal{P}}_j\boldsymbol{\mathcal{X}}_i^\dagger$, the eigenvector $\boldsymbol{u}_{i,j}$ associated with the minimum eigenvalue, $\lambda_{i,j} = \lambda_{\min}(\boldsymbol{\Omega}_{i,j})$, can yield the gradient in terms of $\boldsymbol{\mathcal{X}}_m$ as follows

$$\begin{aligned}
\boldsymbol{\nabla}^{[m]}_{\lambda_{i,j}} &= \frac{\partial \lambda_{i,j}}{\partial \boldsymbol{\mathcal{X}}_m^*} = \frac{\partial}{\partial \boldsymbol{\mathcal{X}}_m^*}\boldsymbol{u}_{i,j}^\dagger\boldsymbol{\Omega}_{i,j}\boldsymbol{u}_{i,j} \\
&= \boldsymbol{u}_{i,j}\boldsymbol{u}_{i,j}^\dagger\boldsymbol{\mathcal{X}}_i\boldsymbol{\mathcal{P}}_j\delta_{i,m} \\
&\quad - \left(\boldsymbol{\mathcal{X}}_j\boldsymbol{\mathcal{X}}_j^\dagger\right)^{-1}\boldsymbol{\mathcal{X}}_j\boldsymbol{\mathcal{X}}_i^\dagger\boldsymbol{u}_{i,j}\boldsymbol{u}_{i,j}^\dagger\boldsymbol{\mathcal{X}}_i\boldsymbol{\mathcal{P}}_j\delta_{j,m}. \qquad (19)
\end{aligned}$$

Here, $\delta_{i,j} = 1$ if $i = j$, otherwise $\delta_{i,j} = 0$. The codebook design using the gradient method is described below:

1: Generate random codewords $\boldsymbol{\mathcal{X}}_m$ such that $\|\boldsymbol{\mathcal{X}}_m\|^2 = M$
2: Compute $\boldsymbol{\Omega}_{i,j}$ to get minimum eigenvalue $\lambda_{i,j} = \lambda_{\min}(\boldsymbol{\Omega}_{i,j})$ for all pairs $i \neq j$
3: Search for the worst pair which has the minimum $\lambda_{i,j}$
4: Compute the eigenvalue $\boldsymbol{u}_{i,j}$ for the worst pair (i, j)
5: Calculate the gradient $\boldsymbol{\nabla}^{[m]}_{\lambda_{i,j}}$ for the pair $m \in \{i, j\}$
6: Update codewords as $\boldsymbol{\mathcal{X}}_m \leftarrow \boldsymbol{\mathcal{X}}_m + \beta\boldsymbol{\nabla}^{[m]}_{\lambda_{i,j}}$, where $\beta \in \mathbb{R}$ is a stepsize factor which is optimized by line searching to maximize $\min_{i \neq j} \lambda_{i,j}$
7: Normalize the energy such that $\|\boldsymbol{\mathcal{X}}_m\|^2 = M$
8: Repeat from 2 until $\lambda_{i,j}$ sufficiently converges

Using multiple initial codewords or small perturbations of optimized codebook, the gradient method yields well–designed codebook. It is also straightforward to adopt this design method for the high–order super–block GLRT because we have $\partial\boldsymbol{\mathcal{X}}'_m = \boldsymbol{\mathcal{D}}\partial\boldsymbol{\Gamma}_m$.

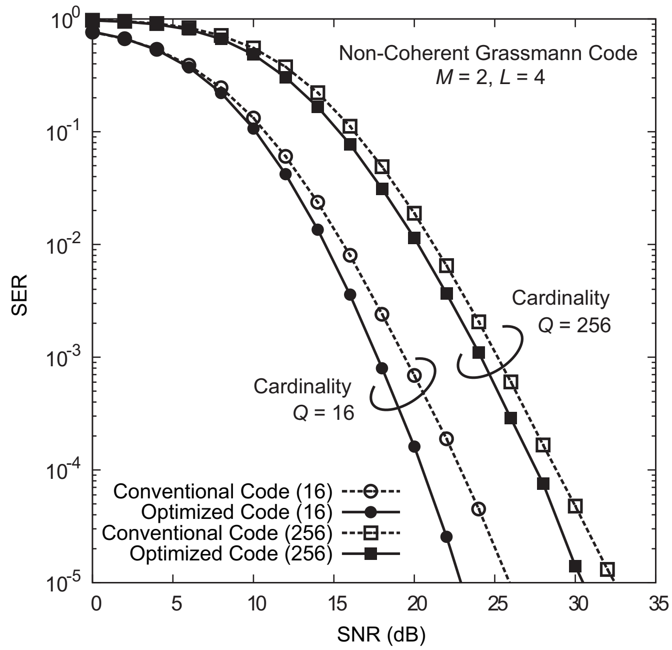

Fig. 1. SER versus SNR for 2×2 MIMO in frequency–flat Rayleigh fading channels ($M = 2$, $L = 4$, $Q \in \{16, 256\}$).

C. Optimization for Exponential Mapping Grassmann Codes

The numerical optimization method was originally used for non–parametric code design in [11]. It is also applicable to the design of parametric non–coherent codes. As an example to show the gains achieved by our optimization method, an improved version of the exponential mapping Grassmann codes is presented here. The conventional Grassmann codes with exponential mapping uses the fixed parameters ϑ and ϕ in (9). We optimize those parameters by the gradient method with a slight modification as

$$\frac{\partial \lambda_{i,j}}{\partial \gamma} = \sum_{m=1}^{Q} \mathrm{tr}\Big[\frac{\partial \boldsymbol{\mathcal{X}}_m^\dagger}{\partial \gamma}\boldsymbol{\nabla}^{[m]}_{\lambda_{i,j}}\Big], \qquad (20)$$

where $\gamma \in \{\alpha, \vartheta, \phi\}$ is a parameter to be optimized.

In Fig. 1, we plot the symbol error rate (SER) performance to compare between the conventional Grassmann codes and the improved Grassmann codes with the optimized parameters, for $M = 2$, $N = 2$ and $L = 4$ over frequency–flat fading channels. One can see from this figure that our design

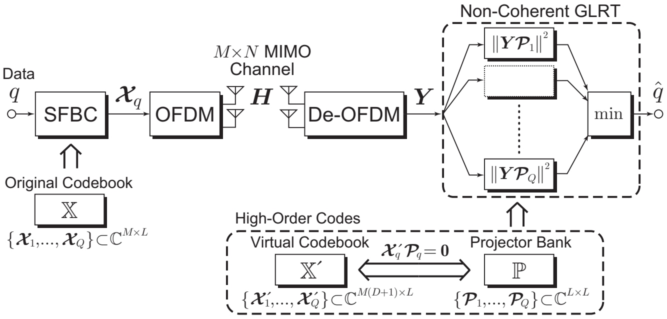

Fig. 2. Non–coherent $M \times N$ MIMO–OFDM systems employing $(M \times L)$–dimensional Grassmann SFBC and D–th order GLRT algorithm.

optimization method offers approximately $3.0\,\mathrm{dB}$ and $2.0\,\mathrm{dB}$ gains for $Q = 16$ and $Q = 256$ cardinalities at an SER of 10^{-5}.

V. Performance Evaluation

Now, we show the performance advantage of our proposed GLRT over the conventional GLRT through computer simulations. Fig. 2 shows an overview of the high–order GLRT algorithm for non–coherent MIMO–OFDM systems. We use $L = 8$ symbols for non–coherent Grassmann codes with a cardinality of $Q = 4$. The transmitter uses $M = 2$ antennas and the receiver uses $N = 2$ antennas. We assume that the channel is frequency–selective Rayleigh fading with an equal–gain power delay profile. The number of subcarriers is 128 so that 16 SFBC blocks are multiplexed over frequency domain via OFDM one symbol. For super–block GLRT, we choose $K = 2$ blocks. For high–order GLRT, the polynomial order is set to be $D \in \{0, 1, 2, 3, 4\}$. The virtual codebook and the projector bank are predetermined by the original codebook.

Figs. 3, 4 and 5 show SER as a function of average SNR for 2–path, 8–path and 16–path fading, respectively as a low–, moderate–, and high–selective channels. The channel correlation factors between adjacent SFBC blocks are 0.98, 0.64 and 0.00 respectively for 2–path, 8–path and 16–path channel conditions. For 2–path channels in Fig. 3, there is a small fluctuation of channel coefficients over subcarriers, that makes the 0–th order super–block GLRT offer the best performance. We can observe more than $3\,\mathrm{dB}$ gains at a SER of 10^{-4} when the sequential super–block GLRT is employed. It is because the high channel correlation between neighboring SFBC blocks can be fully exploited by longer effective code lengths. Hence, further improvement can be expected by using larger K with a cost of computational complexity. The selective decision has $1\,\mathrm{dB}$ inferior performance over the sequential decision, while it still offers better performance than the conventional GLRT by $2\,\mathrm{dB}$. For such a low selective fading, high–order GLRT is of no use because higher–order LS regressions more consume the degrees of freedom for curve fitting (i.e., over–fitting loss).

For 8–path fading channels in Fig. 4, the conventional GLRT (0–th order) has a poor performance which saturates to a SER of 10^{-3} even in the high SNR regimes because the channel frequently changes during one SFBC block due to the low channel correlation of 0.68. It is seen that the 1–st order GLRT has a significant performance improvement in particular for high SNRs. It is remarkable that the computational complexity of higher–order GLRT is exactly same as that of the conventional GLRT. The 2–nd order GLRT degrades its performance gain because the channel selectivity is not so severe. The 2–nd order super–block GLRT (employing the sequential decision) provides additional $4\,\mathrm{dB}$ gains at a SER of 10^{-4} from the 1–st order GLRT. The 1–st order super–block GLRT has a saturating performance like the conventional GLRT. It suggests that we need to use appropriate order D and super–block length K according to the channel condition.

For 16–path channels in Fig. 5, the conventional GLRT does not work (SER performance is much higher than 10^{-1}) because of the highly selective channel over subcarriers (zero channel correlation between adjacent SFBCs). The 4–th order super–block GLRT offers a significant performance improvement over the conventional GLRT receiver. It is surprising that the performance degradation from Fig. 3 can be effectively compensated by high–order super–block GLRT even for the extremely selective 16–path channels which experience zero channel correlation between adjacent SFBC blocks.

Through this section, we demonstrated that our proposed scheme is very advantageous to deal with channel selectivity when we properly choose the order D and the super–block length K. It is highly expected that our proposed scheme still performs well for any practical channel models. It remains some future works regarding the adaptive selection of D and K according to the channel conditions.

VI. Conclusion

In this paper, we propose an improved GLRT receiver termed high–order super–block GLRT for non–coherent MIMO systems in which CSI is not available at either transmitter or receiver. By introducing the high–order LS regressions, our proposed method can deal with a drawback of the conventional GLRT, more specifically, from a severe performance degradation when the channel changes fast during a non–coherent code block. We demonstrated that the high–order super–block GLRT offers a significant performance improvement over the conventional GLRT in particular for highly selective fading channels. There is no increase in computational complexity for the high–order technique while the super–block technique increases the complexity.

In addition to the algorithm development, we optimize non–coherent Grassmann codes to minimize the pairwise error probability through the use of the gradient method as a practical solution for packing spheres on the Grassmann manifold. We exemplified the advantage of our design method by improving the existing Grassmann codes based on exponential mapping. An adaptive selection of the order and super–block parameters remains a subject of further study.

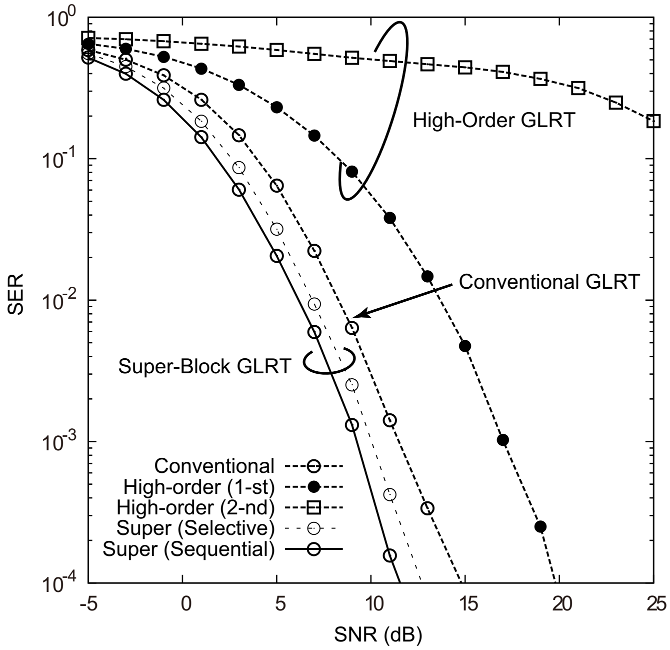

Fig. 3. SER versus SNR for 2×2 MIMO–OFDM in frequency–selective 2–path Rayleigh fading (0.98 channel correlation between adjacent SFBCs).

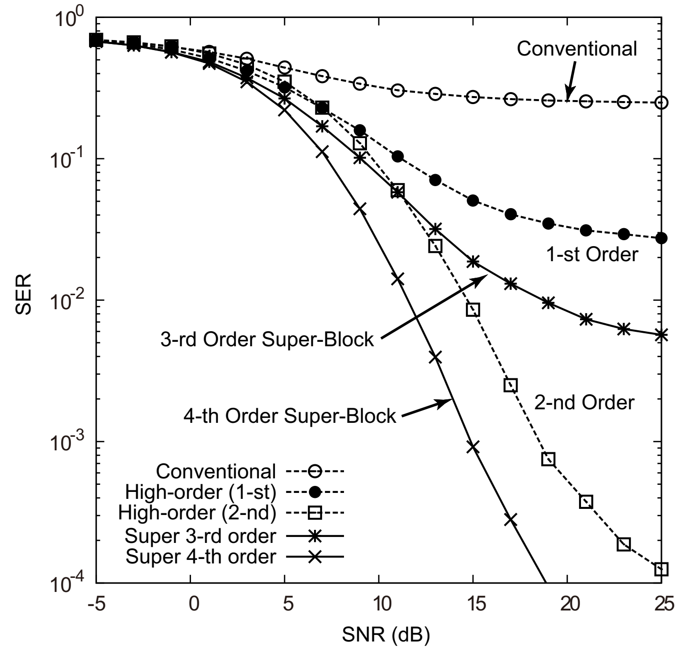

Fig. 5. SER versus SNR for 2×2 MIMO–OFDM in frequency–selective 16–path Rayleigh fading (zero channel correlation between adjacent SFBCs).

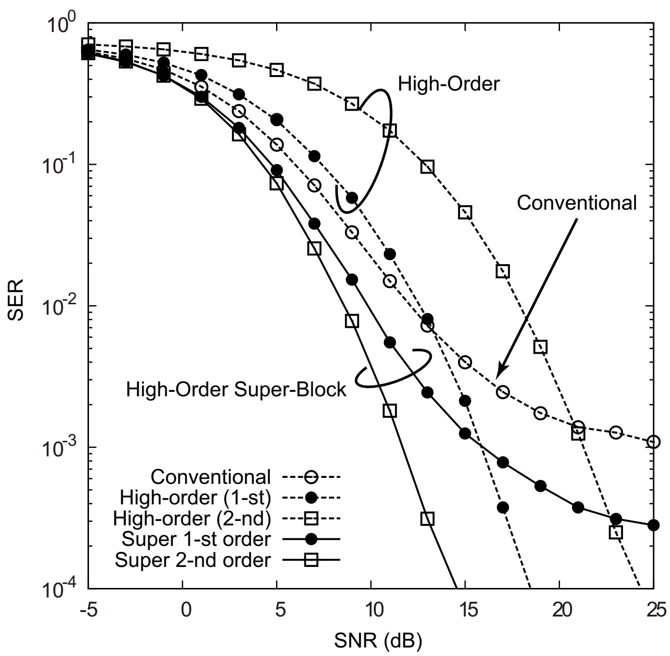

Fig. 4. SER versus SNR for 2×2 MIMO–OFDM in frequency–selective 8–path Rayleigh fading (0.64 channel correlation between adjacent SFBCs).

REFERENCES

[1] G. J. Foschini, “Layered space–time architecture for wireless communication in a fading environment when using multi-element antennas,” *Bell Labs. Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[2] I. E. Telatar, “Capacity of multi-antenna Gaussian channels,” *Eur. Trans. Telecomm.*, vol. 10, no. 6, pp. 585–595, Nov. 1999.

[3] T. Marzetta and B. Hochwald, “Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading,” *IEEE Trans. Inf. Theory*, vol. 45, pp. 139–157, 1999.

[4] L. Zheng and D. N. C. Tse, “Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel,” *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–384, Feb. 2002.

[5] B. M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke, “Systematic design of unitary space–time constellations,” *IEEE Trans. Inf. Theory*, vol. 46, no. 6, Sept. 2000.

[6] Y. Jing and B. Hassibi, “Unitary space-time modulation via Cayley transform,” *IEEE Trans. Signal Proc.*, vol. 51, no. 11, pp. 2891–2904, Nov. 2003.

[7] B. Hochwald and W. Sweldens, “Differential unitary space-time modulation,” *IEEE Trans. Commun.*, vol. 48, pp. 2041–2052, Dec. 2000.

[8] I. Kammoun, A. M. Cipriano, and J.-C. Belfiore, “Non-coherent codes over the Grassmannian,” *IEEE Trans. Wireless Commun.*, vol. 6, no. 10, pp. 3657–3667, Oct. 2007.

[9] I. Kammoun and J. Belfiore, “A new family of Grassmann space-time codes for non-coherent MIMO systems,” *IEEE Commun. Lett.*, vol. 7, no. 11, pp. 528–530, Nov. 2003.

[10] M. J. Borran, A. Sabharwal, and B. Aazhang, “On design criteria and construction of non-coherent space-time constellations,” *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2332–2351, Oct. 2003.

[11] M. Beko, J. Xavier, and V. Barroso, “Codebook design for non-coherent communication in multiple-antenna systems,” *IEEE ICASSP*, Toulouse, France, 2006.

[12] B. L. Hughes, “Differential space–time modulation,” *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2567–2578, Nov. 2000.

[13] V. Tarokh and M. Kim, “Existence and construction of noncoherent unitary space-time codes,” *IEEE Trans. Inf. Theory*, vol. 25, no. 8, pp. 3112–3120, Dec. 2002.

[14] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, New York: Addison-Wesley Publishing Co., 1990.

[15] T. K. Akino, “Optimum-weighted RLS channel estimation for rapid fading MIMO channels,” *IEEE Trans. Wireless Commun.*, vol. 7, no. 11, pp. 4248–4260, Nov. 2008.