SCIENCE CHINA
Information Sciences

xxxx xxxx Vol. xx No. xx: 1–17
doi:

On Channel Quantization for Multi-cell Cooperative Systems with Limited Feedback

Xueying HOU[1]*, Chenyang YANG[1] & Buon Kiong LAU[2]

[1]*School of Electronics and Information Engineering, Beihang University, Beijing* 100191*, China,*
[2]*Department of Electrical and Information Technology, Lund University, Lund,* S-22100*, Sweden.*

Received July 12, 2012; accepted November 20, 2012

Abstract Coherent multi-cell cooperative transmission, also referred to as coordinated multi-point transmission (CoMP), is a promising strategy to provide high spectral efficiency for universal frequency reuse cellular systems. To report the required channel information to the transmitter in frequency division duplexing systems, limited feedback techniques are often applied. Considering that the average channel gains from multiple base stations (BSs) to one mobile station are different and the number of cooperative BSs may be dynamic, it is neither flexible nor compatible to employ a large codebook to directly quantize the CoMP channel. In this paper, we employ per-cell codebooks for quantizing local and cross channels. We first propose a codeword selection criterion, aiming at maximizing an estimated data rate for each user. The proposed criterion can be applied for an arbitrary number of receive antennas at each user and also for an arbitrary number of data streams transmitted to each user. Considering that the resulting optimal per-cell codeword selection for CoMP channel is of high complexity, we propose a serial codeword selection method that has low complexity but yields comparable performance to that of the optimal codeword selection. We evaluate the proposed codeword selection criterion and method using measured CoMP channels from an urban environment as well as simulations. The results demonstrate significant performance gain as compared to an existing low-complexity method.

Keywords Base station cooperative transmission, channel quantization, limited feedback, codeword selection.

1 Introduction

Base station (BS) cooperative transmission, also known as coordinated multi-point transmission (CoMP) in Long Term Evolution Advanced (LTE-A), is an effective way to avoid inter-cell interference in universal frequency reuse cellular systems. CoMP joint processing (CoMP-JP) provides the full benefit of CoMP systems, if both data and channel state information (CSI) can be obtained at a central unit (CU) [1, 2]. For simplicity, we refer to CoMP-JP as CoMP in the following.

CoMP is often viewed as a large multiple-input and multiple-output (MIMO) system with a "super BS" (i.e., the CU). However, there are distinct features in CoMP channels and systems. CoMP channel

*Corresponding author (email: hxymr@ee.buaa.edu.cn). This work was supported in part by the International S&T Cooperation Program of China (ISCP) under Grant No. 2008DFA12100, in part by the Key project of Next Generation Wideband Wireless Communication Network, Ministry of Industry and Information (MII): Coordinated Multiple-point Transmission for IMT-Advanced Systems under Grant No. 2011ZX03003-001, and in part by the Swedish Research Council under Grant No. 2010-468.

is an aggregation of multiple single-cell channels from the cooperative BSs to each user. Considering that the number of cooperating BSs in a cluster may be dynamic, the dimension of the CoMP channel seen by a user may be dynamic. Furthermore, the average channel gains from different BSs to each user are different [2, 3], due to different antenna power gains, path losses and shadowings. As a result, the channels are no longer independent and identically distributed (i.i.d.) and the channel statistics of each user highly depend on its position.

Limited feedback techniques are widely applied for reporting CSI to transmitter in frequency division duplexing (FDD) MIMO systems and have been extensively studied [4]. If conventional methods for single-cell systems are directly applied to design the codebooks for high-dimensional channel matrices in CoMP systems, prohibitive complexity is required to dynamically generate the location-dependent and cluster-dependent codebooks and to search for the optimal codewords. Moreover, frequently re-generating large codebooks is neither flexible nor compatible to existing systems.

In fact, since CoMP channel is an aggregation of multiple single-cell channels, we can reuse the codebook designed for single-cell systems to separately quantize multiple single-cell channels in the global CoMP channel, which is referred to as per-cell codebook quantization [3]. Though this does not yield the optimal codebook for CoMP channel, it can reduce the complexity to generate the codebook as well as the complexity to select the codeword.

In this paper, we study codeword selection for CoMP transmission with per-cell codebook quantization. We first provide a unified codeword selection criterion to maximize an estimated data rate at the user side, which can exploit the feature of CoMP channel, and can accommodate general cases with an arbitrary number of receive antennas at each user and an arbitrary number of data streams transmitted to each user. Codeword selection criteria and methods are well explored for single-cell limited feedback MIMO systems [5, 6, 7, 8, 9]. When each user is equipped with a single antenna and zero-forcing beamforming (ZFBF) is applied, or when each user has multiple antennas and multiple data streams are transmitted to each user with zero-forcing block diagonalization (ZFBD), a widely applied codeword selection criterion is to minimize the chordal distance between the channel direction and the codeword [5, 6]. When multiple antennas are deployed at each user and a single data stream is transmitted to each user, the codeword can be selected with various criteria [7, 8, 9]. It was shown in [8, 9] that the codeword selection jointly designed with a receive combiner outperforms the method of finding the codeword closest to the direction of singular vector corresponding to the maximum singular value of channel matrix. Considering that the selection of per-cell codewords via exhaustive searching is of high complexity [3], we proceed to propose a low complexity method which selects the codewords for per-cell channels in a serial manner. Both simulation results and the results using measured CoMP channels in [10] show that the proposed codeword selection method has minor performance loss from the optimal selection, and outperforms the low-complexity codeword selection method proposed in [3].

To the best of our knowledge, there are few available researches on the codeword selection for CoMP multi-user MIMO (MU-MIMO) systems. A codeword selection method for CoMP MU-MIMO systems with per-cell codebooks was proposed in [3]. Our work differs from that in [3] in three aspects: (1) codeword selection criterion, (2) codeword construction method and (3) codeword selection method to reduce complexity. Due to the first difference, our method can be applied for various numbers of the antennas and data streams at each user, but the method in [3] can only be used when each user has multiple antennas and the received antennas do not provide diversity gain. Due to the second difference, in general cases where the large scale fading gains of a user are different, the proposed method can exploit the difference in the per-cell channel energies to improve the performance of codeword selection. This is because the CoMP channel was normalized by the large scale fading gains of per-cell channels to mimic an i.i.d. single-cell channel in [3]. As a result, the single-cell codeword selection method in [6] can be applied, which selects per-cell codewords by minimizing the chordal distance between the normalized CoMP channel and the aggregated codewords without large scale fading gains [3]. Finally, due to the third difference, we can achieve the same performance as the method proposed in [3] with much lower complexity. Simulation results demonstrate the performance gain of proposed codeword selection criterion and method over that in [3].

Notations: $(x)^*$ and $\mathfrak{Re}(x)$ denote the conjugate and real part of scalar x, respectively. $(\mathbf{X})^T$ and $(\mathbf{X})^H$ denote the transpose, and the conjugate transpose of matrix $\mathbf{X}$, respectively. $\mathrm{tr}\{\mathbf{X}\}$, $\|\mathbf{X}\|_{\mathrm{F}}$, and $\det\{\mathbf{X}\}$ represent the trace, Frobenius norm and determinant of matrix $\mathbf{X}$, respectively. $\mathrm{diag}\{\cdot\}$ is a diagonal matrix. $\mathbb{E}\{\cdot\}$ is the expectation operator. $\mathbf{I}_N$ and $\mathbf{0}_N$ denote an identity matrix of size N and a zero matrix of size N with all elements being 0, respectively. $x \in \mathcal{CN}(\mu, \sigma_x^2)$ represents a random variable x following a complex Gaussian distribution with mean μ and variance σ_x^2. $\triangleq$ denotes a definition operator.

2 System Models

Consider a cellular system with N_B BSs cooperatively serving K mobile stations (MSs). Each BS is equipped with N_T antennas and each user is equipped with N_R antennas. The total number of antennas at all N_B BSs is denoted by $N_T^{\mathrm{sum}} \triangleq N_B N_T$.

The global channel matrix of MS_k is

$$\mathbf{H}_k = [\alpha_{k,1}\mathbf{H}_{k,1}, \ldots, \alpha_{k,N_B}\mathbf{H}_{k,N_B}] = \mathbf{H}_k^w \mathbf{R}_k, \tag{1}$$

where $\alpha_{k,b}$ and $\mathbf{H}_{k,b} \in \mathbb{C}^{N_R \times N_T}$ are respectively the large scale fading gain (including antenna power gain, path loss and shadowing) and the small scale fading channel matrix between BS_b and MS_k, $\mathbf{H}_k^w = [\mathbf{H}_{k,1}, \ldots, \mathbf{H}_{k,N_B}]$ is the aggregated small scale fading channel matrix, and $\mathbf{R}_k = \mathrm{diag}\{\alpha_{k,1}\mathbf{I}_{N_T}, \ldots, \alpha_{k,N_B}\mathbf{I}_{N_T}\}$. It is shown from (1) that the global CoMP channel resembles a special transmit spatially correlated channel. Specifically, the global channel matrix $\mathbf{H}_k$ can be regarded as the transformation of $\mathbf{H}_k^w$ by $\mathbf{R}_k$. To simplify the analysis and highlight the feature of CoMP channels, we assume that the per-cell channels are uncorrelated, and each entry in $\mathbf{H}_{k,b}$ is subject to i.i.d. complex Gaussian random variables with zero mean and unit variance.

We consider linear precoding and denote the precoding matrix of all the cooperative BSs for MS_k by $\mathbf{W}_k \in \mathbb{C}^{N_T^{\mathrm{sum}} \times d_k}$, $k = 1, \ldots, K$, where $d_k \leqslant N_R$ is the number of data streams transmitted to MS_k. Under the assumption of Gaussian transmit signals and additive white Gaussian noise (AWGN), the achievable data rate of MS_k can be expressed as [6]

$$R_k = \log_2 \det\left(\sigma_k^2 \mathbf{I}_{N_R} + \sum_{j=1}^{K} \mathbf{H}_k \mathbf{W}_j \mathbf{W}_j^H \mathbf{H}_k^H\right) - \log_2 \det\left(\sigma_k^2 \mathbf{I}_{N_R} + \sum_{j=1, j\neq k}^{K} \mathbf{H}_k \mathbf{W}_j \mathbf{W}_j^H \mathbf{H}_k^H\right), \tag{2}$$

where σ_k^2 is the variance of each element of the noise vector. To achieve such a data rate, each user only decodes its desired signal and treats co-user interference as noise, and meanwhile, the d_k data streams intended for MS_k are jointly decoded by the maximum likelihood (ML) receiver [11].

2.1 Finite Rate Feedback Model

The required CSI at the BSs for precoding depends on the antenna configuration and the transmission schemes. When the number of receive antennas is equal to the number of data streams and multi-cell ZFBD precoding is applied, the spatial directions of global channel, i.e., the subspace spanned by the columns of $\mathbf{H}_k$, are the required CSI, which need to be quantized and fed back [3, 6]. When multiple antennas are equipped at each MS and only a single data stream is transmitted to each MS by ZFBF, the channel matrix $\mathbf{H}_k$ can be combined into an effective channel vector, which are quantized and fed back to the BSs [8, 9].

In this paper, we consider a unified channel quantization and feedback model, which is applicable for the general case of an arbitrary number of data streams transmitted to each MS. Specifically, we assume that each MS has perfect knowledge of its own global channel matrix. Instead of sending back the full channel knowledge, MS_k can feed back an effective channel matrix $\mathbf{H}_k^{\mathrm{eff}} \triangleq \mathbf{U}_k \mathbf{H}_k \in \mathbb{C}^{d_k \times N_T^{\mathrm{sum}}}$ to reduce the feedback overhead, where $\mathbf{U}_k \in \mathbb{C}^{d_k \times N_R}$ is a combining matrix that converts the global channel matrix with dimensions $N_R \times N_B N_T$ into the effective channel matrix with dimensions $d_k \times N_B N_T$. In order to ensure that the channel vectors in $\mathbf{H}_k^{\mathrm{eff}}$ remains uncorrelated after combining, the combining matrix should be a unitary matrix, i.e., $\mathbf{U}_k \mathbf{U}_k^H = \mathbf{I}_{d_k}$.

Note that the combining matrix $\mathbf{U}_k$ could be applied as the receiver for the desired signal during downlink transmission. When multiple antennas are equipped at each user and only a single data stream is transmitted to each user, i.e., $N_R > 1$ and $d_k = 1$, the combining matrix reduces to a combining vector of size N_R and it can be applied as the receiver [8, 9]. However, it was shown in [9] that such a receiver is inferior to the MMSE receiver designed with the precoded channel $\mathbf{H}_k\mathbf{W}_k$. In this paper, we do not apply the combining matrix as the receiver. As explained earlier, we consider the ML receiver to achieve the data rate shown in (2).

We consider the *per-cell codebook* based limited feedback [3] to quantize $\mathbf{H}_k^{\text{eff}}$. In particular, MS_k employs single-cell codebooks to separately quantize its per-cell effective channels, which are the effective channels from all cooperated BSs, i.e., $\mathbf{H}_{k,b}^{\text{eff}} = \mathbf{U}_k\mathbf{H}_{k,b}, b = 1, \dots, N_B$. We assume that the per-cell large scale channel gains $\alpha_{k,b}$, $b = 1, \dots, N_B$, can be obtained at MS_k by averaging over the received signals and be fed back to the BSs with negligible overhead. After MS_k quantizes each effective per-cell channel matrix $\mathbf{H}_{k,b}^{\text{eff}}$, it feeds back their quantized version to its local BS, i.e., BS_{b_k}, whose received signal has the strongest energy. The cooperative BSs forward their gathered CSI to the CU, who finally reconstructs the global channels for all MSs.

Denote the per-cell codebook for quantizing the effective channel matrix between MS_k and BS_b by $\mathcal{C}_{k,b}$, which consists of $2^{B_{k,b}}$ matrices in $\mathbb{C}^{d_k \times N_T}$, i.e., $\mathbf{V}_{k,b}(1), \dots, \mathbf{V}_{k,b}(2^{B_{k,b}})$, where $B_{k,b}$ is the number of feedback bits allocated to quantize $\mathbf{H}_{k,b}^{\text{eff}}$. For backward compatibility, we consider that the per-cell codewords are unitary matrices [3, 6], i.e., $\mathbf{V}_{k,b}(j)\mathbf{V}_{k,b}^H(j) = \mathbf{I}_{d_k}$. Define the aggregated codeword for the global channel of MS_k as

$$\mathbf{V}_k(i_{k,1}, \dots, i_{k,N_B}) = [\alpha_{k,1}\mathbf{V}_{k,1}(i_{k,1}), \dots, \alpha_{k,N_B}\mathbf{V}_{k,N_B}(i_{k,N_B})] = \mathbf{V}_k^w(i_{k,1}, \dots, i_{k,N_B})\mathbf{R}_k, \tag{3}$$

where $\mathbf{V}_{k,b}(i_{k,b}) \in \mathcal{C}_{k,b}$, $b = 1, \dots, N_B$, $\mathbf{V}_k^w(i_{k,1}, \dots, i_{k,N_B}) = [\mathbf{V}_{k,1}(i_{k,1}), \dots, \mathbf{V}_{k,N_B}(i_{k,N_B})]$ is the aggregated codeword without large scale fading gains, which is the codeword for the aggregated small scale fading channel matrix $\mathbf{H}_k^w$ in (1). Analogous to the special transmit spatially correlated channel structure shown in (1), the aggregated codeword for global channel can be viewed as a transformation of $\mathbf{V}_k^w(i_{k,1}, \dots, i_{k,N_B})$ by $\mathbf{R}_k$.

The channel quantization of MS_k is to find N_B codewords indices, i.e., $\{i_{k,1}^\star, \dots, i_{k,N_B}^\star\}$, in the N_B per-cell codebooks of MS_k, i.e., $\mathcal{C}_{k,1}, \dots, \mathcal{C}_{k,N_B}$, according to some criterion, as will be addressed in Section 3. After MS_k quantizes the effective channel matrices of all the per-cell channels, it feeds back the codeword indices to its local BS, which requires $B_{k,\text{sum}} = \sum_{b=1}^{N_B} B_{k,b}$ bits in total. Then all BSs send the channel information to the CU, and the CU reconstructs the global channel of MS_k as

$$\hat{\mathbf{H}}_k^{\text{eff}} = \mathbf{V}_k(i_{k,1}^\star, \dots, i_{k,N_B}^\star) = [\alpha_{k,1}\mathbf{V}_{k,1}(i_{k,1}^\star), \dots, \alpha_{k,N_B}\mathbf{V}_{k,N_B}(i_{k,N_B}^\star)]. \tag{4}$$

2.2 Multi-cell Scheduling and Precoding

With the reconstructed global channels of all MSs, the CU selects M MSs to serve in the same time-frequency resource with multi-cell ZFBD precoding. ZFBD is a linear precoder for downlink MU-MIMO systems, which has been extensively studied for single-cell transmission [12]. In the special case of multiple-input and single-output (MISO) broadcasting channel, ZFBD reduces to the well-known ZFBF. A major difference between multi-cell ZFBD and single-cell ZFBD lies in the power constraint [13, 14]. While single-cell ZFBD has a sum power constraint (SPC), multi-cell ZFBD should be designed with per-BS power constraint (PBPC). Considering that the optimal ZFBD precoder with PBPC is of high complexity for practical application [14], herein we consider a sub-optimal precoder proposed in [13]. In particular, the quantized channel matrices of all MSs are treated as the true channels and the precoding matrix of MS_k is obtained as

$$\mathbf{W}_k = \mathbf{B}_k\mathbf{M}_k\mathbf{\Lambda}_k^{\frac{1}{2}}, \tag{5}$$

where $\mathbf{B}_k \in \mathbb{C}^{N_T^{\text{sum}} \times (N_T^{\text{sum}} - \sum_{j=1, j\neq k}^K d_j)}$ is the orthonormal basis of the right null space of the matrix formed by stacking all $\hat{\mathbf{H}}_j^{\text{eff}}$, $\forall j \neq k$, together. Specifically, define the effective quantized channel matrix of all MSs other than MS_k as $\hat{\mathbf{H}}_{-k}^{\text{eff}} = [\hat{\mathbf{H}}_1^{\text{eff}\,H}, \dots, \hat{\mathbf{H}}_{k-1}^{\text{eff}\,H}, \hat{\mathbf{H}}_{k+1}^{\text{eff}\,H}, \dots, \hat{\mathbf{H}}_K^{\text{eff}\,H}]^H \in \mathbb{C}^{(\sum_{j=1, j\neq k}^K d_j) \times N_T^{\text{sum}}}$. Then

$\mathbf{B}_k$ is constructed by the last $(N_T^{\rm sum} - \sum_{j=1,j\neq k}^K d_j)$ column vectors of the right-singular matrix of $\hat{\mathbf{H}}_{-k}^{\rm eff}$. $\mathbf{M}_k \in \mathbb{C}^{(N_T^{\rm sum} - \sum_{j=1,j\neq k}^K d_j)\times d_k}$ is the matrix formed by the first d_k column vectors of the right singular matrix of $\hat{\mathbf{H}}_k^{\rm eff}\mathbf{B}_k$, and $\mathbf{\Lambda}_k \in \mathbb{C}^{d_k\times d_k}$ is the diagonal power allocation matrix of MS_k.

The sum-rate maximizing power allocation with PBPC can be founded numerically by convex optimization tools [13], whose complexity is too high for practical use. In this paper, we consider equal power allocation, which is suboptimal but more practical. We consider that the transmit powers of all BSs are the same, which is denoted by P_0. To meet PBPC, the transmit power of all users are scaled by a factor μ as suggested in [13]. Then the power allocation matrix becomes $\mathbf{\Lambda}_k = \mu\frac{N_B P_0}{\sum_{j=1}^K d_j}\mathbf{I}_{d_k}$, where the scaling factor $\mu \in (0,1)$ is given by $\mu = \min\limits_{b=1,\ldots,N_B} \frac{\sum_{j=1}^K d_j/N_B}{\|\mathbf{C}_b\sum_{j=1}^K \mathbf{B}_j\mathbf{M}_j\|_{\rm F}^2}$, $\mathbf{C}_b$ is a block-diagonal matrix of dimension $N_T^{\rm sum}\times N_T^{\rm sum}$ with block size N_T, and the b-th block is $\mathbf{I}_{N_T}$ and other blocks are zeros, $b = 1,\ldots,N_B$.

3 Codeword Selection Criterion

The optimal codeword selection should maximize the achievable data rate of MS_k shown in (2). Nonetheless, the actual data rate of MS_k achieved during data transmission is a function of the precoding matrices of all MSs. When each MS quantizes its own channel, it is unable to know the precoding matrices in advance. To circumvent this problem, we select the codewords to maximize an estimated data rate. In this section, we first propose a codeword selection criterion to accommodate the transmission of an arbitrary number of data streams to each user, and then provide its special forms under various system configurations. Finally, we show the connection of the proposed criterion with an existing one for CoMP systems.

3.1 Proposed Codeword Selection Criterion

When MS_k quantizes its channel, it has neither *a priori* knowledge of the number of MSs scheduled with itself nor the number of data streams transmitted to other MSs. Moreover, it does not know the channels of its own co-scheduled MSs. Therefore, it is impossible for MS_k to know the precoders of all MSs during downlink transmission, which determines the achievable data rate. This is a fundamental challenge in the design of MU-MIMO limited feedback systems. Herein we propose a codeword selection criterion to maximize an estimated data rate of MS_k.

To estimate the downlink data rate, MS_k makes the following three assumptions.

Firstly, full multiplexing is assumed, e.g., $\sum_{j=1}^K d_j = N_T^{\rm sum}$. With this assumption, the matrix $\mathbf{M}_k$ in (5) becomes a unitary matrix of dimension $d_k\times d_k$, which indicates $\mathbf{M}_k\mathbf{M}_k^H = \mathbf{I}_{d_k}$. Secondly, the PBPC is relaxed to SPC, such that the power scaling factor $\mu = 1$. Together with the first assumption, the power allocation matrix in (5) becomes $\mathbf{\Lambda}_k = \frac{N_B P_0}{N_T^{\rm sum}}\mathbf{I}_{d_k}$. Then the term $\mathbf{W}_k\mathbf{W}_k^H$ in (2) can be expressed as

$$\mathbf{W}_k\mathbf{W}_k^H = \mathbf{B}_k\mathbf{M}_k\mathbf{\Lambda}_k\mathbf{M}_k^H\mathbf{B}_k^H = \frac{N_B P_0}{N_T^{\rm sum}}\mathbf{B}_k\mathbf{B}_k^H. \tag{6}$$

The term $\mathbf{B}_k$ is formed by the orthonormal basis of the null space of $\hat{\mathbf{H}}_{-k}^{\rm eff}$, which is the matrix stacked by the effective quantized channel matrices of all MSs other than MS_k. Since MS_k does not have *a prior* information of the quantized channel matrices of other MSs, it is unable to know the true value of $\mathbf{B}_k$. Therefore, we need the third assumption: the scheduled MSs are mutually orthogonal in terms of their quantized effective channel matrices, i.e., $\hat{\mathbf{H}}_k^{\rm eff}\hat{\mathbf{H}}_j^{{\rm eff}^H} = \mathbf{0}$, $j = 1,\ldots,K, j\neq k$. Then, the term $\mathbf{W}_k\mathbf{W}_k^H$ only depends on the quantized channel matrix of MS_k. In practical systems, this is a reasonable assumption when the number of candidate users is sufficiently large [15]. In Section 5, we will verify through simulations that the codeword selection criterion based on the orthogonal scheduling assumption still performs fairly well in realistic scenarios without the assumption. In the following, we derive the expression of $\mathbf{W}_k\mathbf{W}_k^H$.

The orthogonal scheduling assumption indicates that $\hat{\mathbf{H}}_k^{\rm eff}$ lies in the null space of $\hat{\mathbf{H}}_{-k}^{\rm eff}$. Since $\mathbf{B}_k$ forms the orthonormal basis of the null space of $\hat{\mathbf{H}}_{-k}^{\rm eff}$ and $\mathbf{B}_k$ is of dimensions $N_T^{\rm sum}\times d_k$ under full multiplexing assumption, we can express the effective channel matrix as

$$\hat{\mathbf{H}}_k^{\text{eff}} = \mathbf{X}_k \mathbf{B}_k^H, \tag{7}$$

where $\mathbf{X}_k \in \mathbb{C}^{d_k \times d_k}$ is a square matrix.

Then we have $\min\{\text{rank}(\mathbf{X}_k), \text{rank}(\mathbf{B}_k)\} \geqslant \text{rank}(\mathbf{X}_k \mathbf{B}_k^H) = \text{rank}(\hat{\mathbf{H}}_k^{\text{eff}})$. Considering the fact that $\text{rank}(\hat{\mathbf{H}}_k^{\text{eff}}) = d_k$ in order to transmit d_k data streams to MS_k, we have $\min\{\text{rank}(\mathbf{X}_k), \text{rank}(\mathbf{B}_k)\} \geqslant d_k$, which indicates $\text{rank}(\mathbf{X}_k) \geqslant d_k$. Together with the fact that $\mathbf{X}_k$ is a squared matrix of dimension d_k, we can obtain $\text{rank}(\mathbf{X}_k) = d_k$, i.e., the matrix $\mathbf{X}_k$ is full rank and invertible. Then we have

$$\hat{\mathbf{H}}_k^{\text{eff}^H} (\hat{\mathbf{H}}_k^{\text{eff}} \hat{\mathbf{H}}_k^{\text{eff}^H})^{-1} \hat{\mathbf{H}}_k^{\text{eff}} = \mathbf{B}_k \mathbf{X}_k^H (\mathbf{X}_k \mathbf{B}_k^H \mathbf{B}_k \mathbf{X}_k^H)^{-1} \mathbf{X}_k \mathbf{B}_k^H = \mathbf{B}_k \mathbf{B}_k^H. \tag{8}$$

Substituting (8) into (6) gives rise to the following expression:

$$\mathbf{W}_k \mathbf{W}_k^H = \frac{N_B P_0}{N_T^{\text{sum}}} \hat{\mathbf{H}}_k^{\text{eff}^H} (\hat{\mathbf{H}}_k^{\text{eff}} \hat{\mathbf{H}}_k^{\text{eff}^H})^{-1} \hat{\mathbf{H}}_k^{\text{eff}}. \tag{9}$$

Again with orthogonal scheduling assumption and (7), we have $\hat{\mathbf{H}}_k^{\text{eff}} \hat{\mathbf{H}}_j^{\text{eff}^H} = \mathbf{X}_k \mathbf{B}_k^H \mathbf{B}_j \mathbf{X}_j^H = \mathbf{0}$. Recall that $\mathbf{X}_k$ has been shown as invertible, we can obtain $\mathbf{B}_k^H \mathbf{B}_j = \mathbf{0}$. Define $\mathbf{B}_{\text{All}} = [\mathbf{B}_1, \ldots, \mathbf{B}_K] \in \mathbb{C}^{N_T^{\text{sum}} \times (\sum_{k=1}^K d_k)}$. Since $\mathbf{B}_k^H \mathbf{B}_k = \mathbf{I}_{d_k}$, we have $\mathbf{B}_{\text{All}}^H \mathbf{B}_{\text{All}} = \mathbf{I}_{\sum_{k=1}^K d_k}$. With this property and under the assumption of full multiplexing, i.e., $\sum_{k=1}^K d_k = N_T^{\text{sum}}$, we can conclude that $\mathbf{B}_{\text{All}}$ is an $N_T^{\text{sum}} \times N_T^{\text{sum}}$ unitary matrix, i.e., $\mathbf{B}_{\text{All}} \mathbf{B}_{\text{All}}^H = \sum_{k=1}^K \mathbf{B}_k \mathbf{B}_k^H = \mathbf{I}_{N_T^{\text{sum}}}$. Further considering (6), the term $\sum_{k=1}^K \mathbf{W}_k \mathbf{W}_k^H$ in (2) can be expressed as

$$\sum_{k=1}^K \mathbf{W}_k \mathbf{W}_k^H = \frac{N_B P_0}{N_T^{\text{sum}}} \sum_{k=1}^K \mathbf{B}_k \mathbf{B}_k^H = \frac{N_B P_0}{N_T^{\text{sum}}} \mathbf{I}_{N_T^{\text{sum}}}. \tag{10}$$

By substituting (9) and (10) into (2), we can derive the estimated data rate of MS_k as

$$\begin{aligned}
\hat{R}_k =& \log_2 \det\Big(\bar{\sigma}_k^2 \mathbf{I}_{N_R} + \mathbf{H}_k \mathbf{H}_k^H\Big) - \log_2 \det\Big[\bar{\sigma}_k^2 \mathbf{I}_{N_R} + \mathbf{H}_k \mathbf{H}_k^H - \mathbf{H}_k \hat{\mathbf{H}}_k^{\text{eff}^H} (\hat{\mathbf{H}}_k^{\text{eff}} \hat{\mathbf{H}}_k^{\text{eff}^H})^{-1} \hat{\mathbf{H}}_k^{\text{eff}} \mathbf{H}_k^H\Big] \\
\overset{(a)}{=}& \log_2 \det\left\{\Big(\bar{\sigma}_k^2 \mathbf{I}_{N_R} + \mathbf{H}_k \mathbf{H}_k^H\Big) \Big[\bar{\sigma}_k^2 \mathbf{I}_{N_R} + \mathbf{H}_k \mathbf{H}_k^H - \mathbf{H}_k \hat{\mathbf{H}}_k^{\text{eff}^H} (\hat{\mathbf{H}}_k^{\text{eff}} \hat{\mathbf{H}}_k^{\text{eff}^H})^{-1} \hat{\mathbf{H}}_k^{\text{eff}} \mathbf{H}_k^H\Big]^{-1}\right\} \\
=& \log_2 \det\left\{\Big[\mathbf{I}_{N_R} - \mathbf{H}_k \hat{\mathbf{H}}_k^{\text{eff}^H} (\hat{\mathbf{H}}_k^{\text{eff}} \hat{\mathbf{H}}_k^{\text{eff}^H})^{-1} \hat{\mathbf{H}}_k^{\text{eff}} \mathbf{H}_k^H (\bar{\sigma}_k^2 \mathbf{I}_{N_R} + \mathbf{H}_k \mathbf{H}_k^H)^{-1}\Big]^{-1}\right\} \\
=& -\log_2 \det\left\{\mathbf{I}_{N_R} - \mathbf{H}_k \hat{\mathbf{H}}_k^{\text{eff}^H} (\hat{\mathbf{H}}_k^{\text{eff}} \hat{\mathbf{H}}_k^{\text{eff}^H})^{-1} \hat{\mathbf{H}}_k^{\text{eff}} \mathbf{H}_k^H (\bar{\sigma}_k^2 \mathbf{I}_{N_R} + \mathbf{H}_k \mathbf{H}_k^H)^{-1}\right\},
\end{aligned} \tag{11}$$

where (a) is derived from two facts: 1) $-\log_2 \det(\mathbf{M}) = \log_2[1/\det(\mathbf{M})] = \log_2 \det(\mathbf{M}^{-1})$ for an arbitrary invertible matrix $\mathbf{M}$; 2) $\det(\mathbf{MN}) = \det(\mathbf{M})\det(\mathbf{N})$ for arbitrary matrices $\mathbf{M}$ and $\mathbf{N}$, and $\bar{\sigma}_k^2 = \frac{\sigma_k^2 N_T^{\text{sum}}}{N_B P_0}$ is the normalized noise variance.

The per-cell codewords can be selected to maximize the estimated data rate. Specifically, we can first calculate $\hat{R}_k$ by setting $\hat{\mathbf{H}}_k^{\text{eff}} = \mathbf{V}_k(i_{k,1}, \ldots, i_{k,N_B}) = [\alpha_{k,1} \mathbf{V}_{k,1}(i_{k,1}), \ldots, \alpha_{k,N_B} \mathbf{V}_{k,N_B}(i_{k,N_B})]$, $\mathbf{V}_{k,b}(i_{k,b}) \in \mathcal{C}_{k,b}$, $b = 1, \ldots, N_B$. Then, we find the per-cell codewords indices $\{i_{k,1}^\star, \ldots, i_{k,N_B}^\star\}$ that maximize $\hat{R}_k$. Because unitary per-cell codewords are applied, we have $\mathbf{V}_k(i_{k,1}, \ldots, i_{k,N_B}) \mathbf{V}_k^H(i_{k,1}, \ldots, i_{k,N_B}) = \sum_{b=1}^{N_B} \alpha_{k,b}^2 \mathbf{V}_{k,b}(i_{k,b}) \mathbf{V}_{k,b}^H(i_{k,b}) = \sum_{b=1}^{N_B} \alpha_{k,b}^2$.

Considering that the selected codewords are the quantized version of the effective channel matrix $\hat{\mathbf{H}}_k^{\text{eff}}$, the optimal combining matrix $\mathbf{U}_k$ is implicitly included in the selected per-cell codewords $\mathbf{V}_{k,b}(i_{k,b}^\star)$, $b = 1, \ldots, N_B$. By observing the dimensions of the matrices $\hat{\mathbf{H}}_k^{\text{eff}}$, $\mathbf{V}_k(i_{k,1}, \ldots, i_{k,N_B})$ and $\mathbf{H}_k$, we can see that the function of the combining matrix $\mathbf{U}_k$ is to reduce the dimension of global channel matrix $\mathbf{H}_k$ before quantization based on the number of data streams d_k. Therefore, it is unnecessary to provide its explicit expression.

The codeword selection problem can be described as the following proposition.

Proposition 1. Finding the codewords indices $\{i^\star_{k,1},\ldots,i^\star_{k,N_B}\}$ for MS_k that maximize the estimated data rate of MS_k can be formulated as the following problem:

$$\begin{aligned}&\min_{i_{k,1},\ldots,i_{k,N_B}} \quad f(\mathbf{V}_k(i_{k,1},\ldots,i_{k,N_B}))\\ &\quad s.t. \quad \mathbf{V}_{k,b}(i_{k,b})\in\mathcal{C}_{k,b},\ \forall b=1,\ldots,N_B,\end{aligned}\tag{12}$$

where the expression of objective function is

$$\begin{aligned}&f(\mathbf{V}_k(i_{k,1},\ldots,i_{k,N_B}))\triangleq\\ &\quad\det\left\{\mathbf{I}_{N_R}-\frac{1}{\sum_{b=1}^{N_B}\alpha_{k,b}^2}\mathbf{H}_k\mathbf{V}_k^H(i_{k,1},\ldots,i_{k,N_B})\mathbf{V}_k(i_{k,1},\ldots,i_{k,N_B})\mathbf{H}_k^H(\bar{\sigma}_k^2\mathbf{I}_{N_R}+\mathbf{H}_k\mathbf{H}_k^H)^{-1}\right\}.\end{aligned}\tag{13}$$

In the following, we show the resulting criteria in the *Proposition* under various system configurations and the connection with existing criteria in literature.

3.1.1 *Criterion under configuration $N_R=1$, $d_k=1$*

When MS_k is equipped with a single antenna, the combining matrix reduces to a scalar, and its downlink global channel degenerates to a vector, i.e., $\mathbf{h}_k\in\mathbb{C}^{1\times N_T^{\mathrm{sum}}}$. The global codeword of MS_k also degenerates to a vector, i.e., $\mathbf{v}_k(i_{k,1},\ldots,i_{k,N_B})=[\alpha_{k,1}\mathbf{v}_{k,1}(i_{k,1}),\ldots,\alpha_{k,N_B}\mathbf{v}_{k,N_B}(i_{k,N_B})]\in\mathbb{C}^{1\times N_T^{\mathrm{sum}}}$. The objective function of codeword selection problem in (13) becomes

$$f(\mathbf{v}_k(i_{k,1},\ldots,i_{k,N_B}))=1-\frac{|\mathbf{v}_k(i_{k,1},\ldots,i_{k,N_B})\mathbf{h}_k^H|^2}{(\sum_{b=1}^{N_B}\alpha_{k,b}^2)(\bar{\sigma}_k^2+\|\mathbf{h}_k\|^2)}.\tag{14}$$

We can verify that finding the per-cell codewords $\mathbf{v}_{k,b}(i_{k,b})\in\mathcal{C}_{k,b},\ \forall b=1,\ldots,N_B$, minimizing (14) is equivalent to minimizing the chordal distance between $\mathbf{v}_k(i_{k,1},\ldots,i_{k,N_B})$ and $\mathbf{h}_k$, whose definition is $d^2(\mathbf{m},\mathbf{n})=1-\frac{|\mathbf{m}^H\mathbf{n}|^2}{\|\mathbf{m}\|^2\|\mathbf{n}\|^2}$ for arbitrary column vectors $\mathbf{m}$ and $\mathbf{n}$ [4].

3.1.2 *Criterion under configuration $N_R>1$, $d_k=1$*

When MS_k has more than one antenna and only a single data stream is transmitted to the MS, its global codeword is a vector, i.e., $\mathbf{v}_k(i_{k,1},\ldots,i_{k,N_B})=[\alpha_{k,1}\mathbf{v}_{k,1}(i_{k,1}),\ldots,\alpha_{k,N_B}\mathbf{v}_{k,N_B}(i_{k,N_B})]\in\mathbb{C}^{1\times N_T^{\mathrm{sum}}}$, and the combining matrix is also a vector of size N_R. Then the objective function in (13) becomes

$$\begin{aligned}&f(\mathbf{v}_k(i_{k,1},\ldots,i_{k,N_B}))\\ &\quad=\det\left\{\mathbf{I}_{N_R}-\frac{1}{\sum_{b=1}^{N_B}\alpha_{k,b}^2}\mathbf{H}_k\mathbf{v}_k^H(i_{k,1},\ldots,i_{k,N_B})\mathbf{v}_k(i_{k,1},\ldots,i_{k,N_B})\mathbf{H}_k^H(\bar{\sigma}_k^2\mathbf{I}_{N_R}+\mathbf{H}_k\mathbf{H}_k^H)^{-1}\right\}\\ &\quad=1-\frac{1}{\sum_{b=1}^{N_B}\alpha_{k,b}^2}\mathbf{v}_k(i_{k,1},\ldots,i_{k,N_B})\mathbf{H}_k^H(\bar{\sigma}_k^2\mathbf{I}_{N_R}+\mathbf{H}_k\mathbf{H}_k^H)^{-1}\mathbf{H}_k\mathbf{v}_k^H(i_{k,1},\ldots,i_{k,N_B}),\end{aligned}\tag{15}$$

where the last step is obtained from the fact that $\det\left\{\mathbf{I}-\mathbf{m}\mathbf{n}^H\right\}=1-\mathbf{m}^H\mathbf{n}$ for arbitrary column vectors $\mathbf{m}$ and $\mathbf{n}$.

We can verify that selecting codewords to minimize (15) is the same as the codeword selection criterion proposed in [9] for single-cell MU-MIMO systems, which is derived by combining the received signals at multiple antennas of each MS to maximize the expected signal-to-interference-plus-noise ratio (SINR). As shown in [9], this criterion corresponds to the criterion derived by combining the received signals at multiple antennas with the quantization-based combining proposed in [8] when $\bar{\sigma}_k^2$ is small, and reduces to that derived by maximum receive combining [16] when $\bar{\sigma}_k^2$ is large.

3.1.3 *Criterion under configuration $N_R>1$, $d_k=N_R$*

When the number of data streams transmitted to MS_k is N_R, its global codeword $\mathbf{V}_k^H(i_{k,1},\ldots,i_{k,N_B})$ becomes a matrix of size $N_R\times N_T^{\mathrm{sum}}$. The objective function in (13) can be approximated as

$$f(\mathbf{V}_k(i_{k,1},\ldots,i_{k,N_B})) \approx$$
$$1-\frac{1}{\sum_{b=1}^{N_B}\alpha_{k,b}^2}\mathrm{tr}\Big\{\mathbf{H}_k\mathbf{V}_k^H(i_{k,1},\ldots,i_{k,N_B})\mathbf{V}_k(i_{k,1},\ldots,i_{k,N_B})\mathbf{H}_k^H(\bar{\sigma}_k^2\mathbf{I}_{N_R}+\mathbf{H}_k\mathbf{H}_k^H)^{-1}\Big\}, \quad (16)$$

where the approximation comes by considering that $\det(\mathbf{I}+\epsilon\mathbf{M})\approx 1+\epsilon\mathrm{tr}(\mathbf{M})$ when the constant ϵ is small.

When the SNR is high, i.e., $\bar{\sigma}_k^2 \ll \|\mathbf{H}_k\mathbf{H}_k^H\|_{\mathrm{F}}^2$, (16) becomes

$$f_{\mathrm{HSNR}}^{\mathrm{app}}(\mathbf{V}_k(i_{k,1},\ldots,i_{k,N_B}))$$
$$\triangleq 1-\frac{1}{\sum_{b=1}^{N_B}\alpha_{k,b}^2}\mathrm{tr}\{\mathbf{V}_k^H(i_{k,1},\ldots,i_{k,N_B})\mathbf{H}_k^H(\mathbf{H}_k\mathbf{H}_k^H)^{-1}\mathbf{H}_k\mathbf{V}_k(i_{k,1},\ldots,i_{k,N_B})\}. \quad (17)$$

When the SNR is low, i.e., $\bar{\sigma}_k^2 \gg \|\mathbf{H}_k\mathbf{H}_k^H\|_{\mathrm{F}}^2$, (16) turns into

$$f_{\mathrm{LSNR}}^{\mathrm{app}}(\mathbf{V}_k(i_{k,1},\ldots,i_{k,N_B})) \triangleq 1-\frac{1}{\bar{\sigma}_k^2\sum_{b=1}^{N_B}\alpha_{k,b}^2}\|\mathbf{H}_k\mathbf{V}_k^H(i_{k,1},\ldots,i_{k,N_B})\|_{\mathrm{F}}^2. \quad (18)$$

It is easy to verify that selecting the per-cell codewords by minimizing (17) corresponds to minimizing the chordal distance between $\mathbf{V}_k(i_{k,1},\ldots,i_{k,N_B})$ and $\mathbf{H}_k$. For matrices $\mathbf{M}$ and $\mathbf{N}$ of size $N_r \times N_c$ and $N_r \leqslant N_c$, the chordal distance is defined as $d^2(\mathbf{M},\mathbf{N}) = N_r - \mathrm{tr}\{\mathbf{M}^H(\mathbf{M}\mathbf{M}^H)^{-1}\mathbf{M}\mathbf{N}^H(\mathbf{N}\mathbf{N}^H)^{-1}\mathbf{N}\}$ [6]. Meanwhile, minimizing (18) is the same as maximizing the data rate of MS_k under single-user transmission, which was proposed in [17].

3.2 Relationship with an Existing Codeword Selection Criterion for CoMP systems

In [3], a per-cell codebook based limited feedback scheme was proposed for the case where $d_k = N_R$. Remind that we have employed a "transformed" global codeword to quantize the CoMP channel direction to incorporate the channel imbalance feature of CoMP channel, as shown in (3). By contrast, the method in [3] converts CoMP channels to i.i.d. channels in order to apply the codeword selection methods for single-cell systems. Specifically, the authors in [3] selected the per-cell codewords aiming at minimizing the chordal distance between the aggregated small scale fading channel $\mathbf{H}_k^w$ shown in (1) and the aggregated small scale fading codeword $\mathbf{V}_k^w(i_{k,1},\ldots,i_{k,N_B})$ shown in (3). The codeword selection problem was described as the following problem in [3]:

$$\begin{aligned}\min_{i_{k,1},\ldots,i_{k,N_B}} \quad & g(\mathbf{V}_k^w(i_{k,1},\ldots,i_{k,N_B}))\\ s.t. \quad & \mathbf{V}_{k,b}(i_{k,b})\in\mathcal{C}_{k,b},\ \forall b=1,\ldots,N_B,\end{aligned} \quad (19)$$

where $g(\mathbf{V}_k^w(i_{k,1},\ldots,i_{k,N_B})) \triangleq N_R - \frac{1}{N_B}\mathrm{tr}\{\mathbf{V}_k^w(i_{k,1},\ldots,i_{k,N_B})\mathbf{H}_k^{w^H}(\mathbf{H}_k^w\mathbf{H}_k^{w^H})^{-1}\mathbf{H}_k^w\mathbf{V}_k^{w^H}(i_{k,1},\ldots,i_{k,N_B})\}$.

After MS_k finds the per-cell codewords from (19), it feeds back the indices of selected codewords, $i_{k,1}^\star,\ldots,i_{k,N_B}^\star$. Then, the CU reconstructs the quantized CoMP channel according to (4), which is $\hat{\mathbf{H}}_k^{\mathrm{eff}} = \mathbf{V}_k(i_{k,1}^\star,\ldots,i_{k,N_B}^\star)$.

Comparing the objective function of problem (19) and $f_{\mathrm{HSNR}}^{\mathrm{app}}(\mathbf{V}_k(i_{k,1},\ldots,i_{k,N_B}))$ in (17), which is the approximation of our objective function at high SNR, we can observe that they are the same only when all the large scale fading gains of MS_k are equal, i.e., $\alpha_{k,1}=\cdots=\alpha_{k,N_B}\triangleq\alpha_{\mathrm{edge}}$. Under this scenario, the global channel of MS_k reduces to $\mathbf{H}_k=\alpha_{\mathrm{edge}}\mathbf{H}_k^w$, and the global channel codeword becomes $\mathbf{V}_k(i_{k,1},\ldots,i_{k,N_B}) = \alpha_{\mathrm{edge}}\mathbf{V}_k^w(i_{k,1},\ldots,i_{k,N_B})$. It is clear that in this case minimizing the objective function of (19) and (17) lead to the same codewords. However, in practical systems when considering the path loss, shadowing and sector antenna power gains, the large scale fading gains of MS_k will be different in a large probability. This indicates that in general, selecting codewords according to (19) does not ensure the minimization of chordal distance between the global channel $\mathbf{H}_k$ and the reconstructed channel $\hat{\mathbf{H}}_k^{\mathrm{eff}}$.

4 Serial Codeword Selection

4.1 Serial Codeword Selection

The problem of (12) is a standard combinatorial optimization problem and the optimal solution requires an exhaustive searching over the N_B per-cell codebooks. Moreover, from the expression of the objective function in (13) we can observe that the operation of matrix determinant is required during the combinatorial search, whose order of complexity is $\mathcal{O}(N_R^3)$. Owing to these two aspects, the complexity of the codeword selection method from solving problem (12) is too high for MS to afford in practice. In the following, based on the observation that the codewords of different per-cell channels have different impacts on the objective function, we propose a low-complexity codeword selection method.

Rather than minimize the objective function shown in (13), which requires matrix determinant operation, we can minimize its approximation shown in (16), which approximates the determinant of matrix by the operation of matrix trace. This will reduce the complexity significantly when the value of N_R is large. After some regular derivations, we can further show that the approximation in (16) is the same as (13) when a single data stream is transmitted to each MS. When multiple data streams are transmitted to each MS, the approximation will lead to a performance loss, which is however not severe, as will be shown in simulation.

Minimizing the approximation of objective function shown in (16) is equivalent to maximizing

$$\begin{aligned}
&\bar{f}^{\text{app}}(\mathbf{V}_k(i_{k,1},\ldots,i_{k,N_B}))\\
&\quad\triangleq \text{tr}\Big\{\mathbf{H}_k\mathbf{V}_k^H(i_{k,1},\ldots,i_{k,N_B})\mathbf{V}_k(i_{k,1},\ldots,i_{k,N_B})\mathbf{H}_k^H(\bar{\sigma}_k^2\mathbf{I}_{N_R}+\mathbf{H}_k\mathbf{H}_k^H)^{-1}\Big\}\\
&\quad\overset{(a)}{=}\text{tr}\left\{(\bar{\sigma}_k^2\mathbf{I}_{N_R}+\mathbf{H}_k\mathbf{H}_k^H)^{-1}\sum_{b=1}^{N_B}\alpha_{k,b}^2\mathbf{H}_{k,b}\mathbf{V}_{k,b}^H(i_{k,b})\sum_{a=1}^{N_B}\alpha_{k,a}^2\mathbf{V}_{k,a}(i_{k,a})\mathbf{H}_{k,a}^H\right\}\\
&\quad=\sum_{b=1}^{N_B}\sum_{a=1}^{N_B}\alpha_{k,b}^2\alpha_{k,a}^2\underbrace{\text{tr}\left\{(\bar{\sigma}_k^2\mathbf{I}_{N_R}+\mathbf{H}_k\mathbf{H}_k^H)^{-1}\mathbf{H}_{k,b}\mathbf{V}_{k,b}^H(i_{k,b})\mathbf{V}_{k,a}(i_{k,a})\mathbf{H}_{k,a}^H\right\}}_{\beta_k(b,a)},
\end{aligned}\tag{20}$$

where (a) is obtained by substituting the expressions of global channel matrix and the global codeword of MS_k shown in (1) and (3).

We can observe that $\bar{f}^{\text{app}}(\mathbf{V}_k(i_{k,1},\ldots,i_{k,N_B}))=(\bar{f}^{\text{app}}(\mathbf{V}_k(i_{k,1},\ldots,i_{k,N_B})))^*$, which implies that it is a real scalar and its expression can be further derived as

$$\begin{aligned}
\bar{f}^{\text{app}}(\mathbf{V}_k(i_{k,1},\ldots,i_{k,N_B}))&=\mathfrak{Re}\left\{\bar{f}^{\text{app}}(\mathbf{V}_k(i_{k,1},\ldots,i_{k,N_B}))\right\}\\
&=\sum_{b=1}^{N_B}\alpha_{k,b}^4\beta_k(b,b)+\sum_{b=1}^{N_B}\sum_{a=1,a\neq b}^{N_B}\alpha_{k,b}^2\alpha_{k,a}^2\mathfrak{Re}\left\{\beta_k(b,a)\right\} \qquad (21)\\
&\overset{(a)}{=}\sum_{b=1}^{N_B}\alpha_{k,b}^4\beta_k(b,b)+\sum_{b=1}^{N_B}\sum_{a=1}^{b-1}2\alpha_{k,b}^2\alpha_{k,a}^2\mathfrak{Re}\left\{\beta_k(b,a)\right\}\\
&=\sum_{b=1}^{N_B}\alpha_{k,b}^2\underbrace{\left[\alpha_{k,b}^2\beta_k(b,b)+\sum_{a=1}^{b-1}2\alpha_{k,a}^2\mathfrak{Re}\left\{\beta_k(b,a)\right\}\right]}_{\Gamma_{k,b}}, \qquad (22)
\end{aligned}$$

where (a) is obtained from the fact $\beta_k(b,a)=(\beta_k(a,b))^*$.

Now we see that the objective function can be expressed as a weighted summation of $\Gamma_{k,b}$ defined in (22), and the weighting coefficients are the squared large scale fading gains of the links between MS_k and BSs. As stated in Section 3, the large scale fading gains of MS_k are different with high probability. Therefore, the values of $\Gamma_{k,b}$ for different b have different contributions to the final objective function. For a strong link, i.e., a large value of $\alpha_{k,b}^2$, the value of $\Gamma_{k,b}$ plays an important role in the objective function. By contrast, for a weak link, i.e., a small value of $\alpha_{k,b}^2$, the value of $\Gamma_{k,b}$ has an insignificant contribution to the objective function.

The expression of $\Gamma_{k,b}$ includes both $\beta_k(b,b) = \mathrm{tr}\left\{(\bar{\sigma}_k^2\mathbf{I}_{N_R} + \mathbf{H}_k\mathbf{H}_k^H)^{-1}\mathbf{H}_{k,b}\mathbf{V}_{k,b}^H(i_{k,b})\mathbf{V}_{k,b}(i_{k,b})\mathbf{H}_{k,b}^H\right\}$ and $\mathfrak{Re}\{\beta_k(b,a)\} = \mathfrak{Re}\left\{\mathrm{tr}\left\{(\bar{\sigma}_k^2\mathbf{I}_{N_R} + \mathbf{H}_k\mathbf{H}_k^H)^{-1}\mathbf{H}_{k,b}\mathbf{V}_{k,b}^H(i_{k,b})\mathbf{V}_{k,a}(i_{k,a})\mathbf{H}_{k,a}^H\right\}\right\}$, $a = 1,\ldots,b-1$. The value of $\beta_k(b,b)$ is determined only by the index of codeword $\mathbf{V}_{k,b}(i_{k,b})$, and can be considered as the individual part of the per-cell codeword. The value of $\mathfrak{Re}\{\beta_k(b,a)\}$ depends on the indices of both $\mathbf{V}_{k,b}(i_{k,b})$ and $\mathbf{V}_{k,b}(i_{k,a})$, and can be considered as the interactive part of two per-cell codewords. As shown in (22), when selecting the codeword index of $\mathbf{V}_{k,b}(i_{k,b})$, both the individual part and the interacting parts related to this codeword should be taken into consideration.

Based on these observations, we propose a serial codeword selection, which is to select the codeword for each per-cell channel matrix in a serial manner, whose order depends on the contribution of $\Gamma_{k,b}$ to the objective function. Specifically, we sort the per-cell channel matrices indices according to the descending order of average gains of per-cell channels, i.e., $\alpha_{k,b}^2$. Define the sorted indices vector as $\Omega \triangleq [l_1,\ldots,l_{N_B}]$, where l_i represents the index of per-cell channel matrix with the ith largest average gain. Considering that the value of Γ_{k,l_1} contributes most to the objective function, we first choose a codeword for this per-cell channel matrix to maximize $\Gamma_{k,l_1} = \alpha_{k,l_1}^2\beta_k(l_1,l_1)$. Next, we quantize the l_2th per-cell channel. If the per-cell channel with the second largest average channel gain is quantized independently, we can obtain a codeword to maximize the individual part related to this codeword, i.e., $\alpha_{k,l_2}^2\beta_k(l_2,l_2)$. However, this does not ensure the maximization of the interacting part $2\alpha_{k,l_1}^2\mathfrak{Re}\{\beta_k(l_1,l_2)\}$, whose value depends on the per-cell channel matrices and codewords of both l_1th and l_2th per-cell channels. Therefore, when selecting the codeword for the l_2th per-cell channel, we should choose a codeword from codebook $\mathcal{C}_{k,l_2}$ that maximizes $\Gamma_{k,l_2} = \alpha_{k,l_2}^2\beta_k(l_2,l_2) + 2\alpha_{k,l_1}^2\mathfrak{Re}\{\beta_k(l_2,l_1)\}$.

The procedure of the serial codeword selection method is summarized as follows.

Serial Per-cell Codeword Selection

Step 1: Sort the per-cell channel matrices indices in descending order of their large scale fading gains $\alpha_{k,b}^2$ as $\Omega = [l_1,\ldots,l_{N_B}]$.

Step 2: Initialize the codeword selection by setting $j = 1$.

Step 3: Choose the quantization of effective channel matrix with the jth largest large scale fading gain as $\mathbf{V}_{k,l_j}(i_{k,l_j}^\star)$, whose index is chosen as

$$i_{k,l_j}^\star = \mathop{\arg\max}_{\mathbf{V}_{k,l_j}(i_{k,l_j})\in\mathcal{C}_{k,l_j}} \left(\alpha_{k,l_j}^2\beta_k(l_j,l_j) + \sum_{b=1}^{j-1} 2\alpha_{k,l_b}^2\mathfrak{Re}\{\beta_k(l_j,l_b)\}\right),$$

where $\beta_k(l_j,l_j) = \mathrm{tr}\left\{(\bar{\sigma}_k^2\mathbf{I}_{N_R} + \mathbf{H}_k\mathbf{H}_k^H)^{-1}\mathbf{H}_{k,l_j}\mathbf{V}_{k,l_j}^H(i_{k,l_j})\mathbf{V}_{k,l_j}(i_{k,l_j})\mathbf{H}_{k,l_j}^H\right\}$, $\beta_k(l_j,l_b) = \mathrm{tr}\left\{(\bar{\sigma}_k^2\mathbf{I}_{N_R} + \mathbf{H}_k\mathbf{H}_k^H)^{-1}\mathbf{H}_{k,l_j}\mathbf{V}_{k,l_j}^H(i_{k,l_j})\mathbf{V}_{k,l_b}(i_{k,l_b}^\star)\mathbf{H}_{k,l_b}^H\right\}$, and $\mathbf{V}_{k,l_b}(i_{k,l_b}^\star)$ is the selected codeword for the l_bth channel in the previous steps.

Step 4: $j = j+1$. If $j \leqslant N_B$, go to step 3, otherwise stop the selection algorithm.

4.2 Complexity Analysis

When the per-cell codewords are selected by maximizing the objective function in (20) through an exhaustive searching, which is referred to as joint codeword selection method in this subsection, we can show that its order of complexity is $\mathcal{O}(\prod_{b=1}^{N_B} 2^{B_{k,b}})$.

A per-cell codeword selection method of low complexity was proposed in [3]. The basic idea is to first construct a sub-codebook with codewords that lie in the neighborhood of the per-cell channel to be quantized, and then to find the indices through exhaustive searching among the reconstructed sub-codebooks to maximize the objective function in (19). The order of complexity of the first step is $\mathcal{O}(\sum_{b=1}^{N_B} 2^{B_{k,b}})$, and the order of complexity of the second step is $\mathcal{O}(\prod_{b=1}^{N_B} \varphi_{k,b})$, where $\varphi_{k,b}$ is the cardinality of the sub-codebook for quantizing the bth per-cell channel. A tradeoff between complexity and performance can be adjusted by the range of the neighborhood, i.e. the size of $\varphi_{k,b}$.

From the procedure of the proposed serial per-cell codeword selection method, we can observe that to quantize the l_jth per-cell channel of MS_k, we only need to search for a codeword in the codebook $\mathcal{C}_{k,l_j}$. Thereby the order of complexity of the l_jth step is $\mathcal{O}(2^{B_{k,l_j}})$. The serial codeword selection includes N_B steps and its overall complexity is on the order of $\mathcal{O}(\sum_{b=1}^{N_B} 2^{B_{k,b}})$. For ease of comparison, the computational complexity of the three codeword selection methods are summarized in Table 1.

Table 1: Computational complexity of three codeword selection methods

Methods	Computational Complexity
Joint Codeword Selection Method	$\mathcal{O}(\prod_{b=1}^{N_B} 2^{B_{k,b}})$
Method in [3]	$\mathcal{O}(\sum_{b=1}^{N_B} 2^{B_{k,b}}) + \mathcal{O}(\prod_{b=1}^{N_B} \varphi_{k,b})$
Serial Codeword Selection Method	$\mathcal{O}(\sum_{b=1}^{N_B} 2^{B_{k,b}})$

As an example, we consider a case where the number of cooperative BSs $N_B = 3$, and MS_k is located at the exact cell edge of the three cells. This setup indicates that the large scale fading gains from the three BSs to MS_k are equal. Let the size of three per-cell codebooks be $B_{k,1} = B_{k,2} = B_{k,3} = 4$ bits. Then, the order of complexity of joint codeword selection is $\mathcal{O}(4096)$. The order of complexity of the codeword selection method in [3] is $\mathcal{O}(48 + \prod_{b=1}^{3} \varphi_{k,b})$. When $\varphi_{k,b} = 8$, which means that the size of sub-codebook is half of the original codebook, the complexity is on the order of $\mathcal{O}(560)$. In contrast, the complexity of the proposed serial codeword selection method is only on the order of $\mathcal{O}(48)$.

5 Simulation Results

In this section, the performance of different codeword selection methods will first be compared via simulation and then using measured channels from an urban environment.

5.1 Performance Comparison with Simulated Topology and Channel Model

5.1.1 Simulation Setup

We consider a CoMP system with three faced sectors forming a cooperative cluster, as shown in Fig. 1. Each BS is equipped with four antennas. The sector antenna power gain is a function of the horizontal angle ϕ (in degrees) follows 3GPP LTE specification [18], i.e., $\mathrm{AG}^{\mathrm{dB}} = 14 - \min\{12(\frac{\phi}{70})^2, 20\}$, $-\pi < \phi < \pi$. The path-loss model is $\mathrm{PL}^{\mathrm{dB}} = 35.3 + 37.6 \log_{10}(d_{k,b})$, which is employed in LTE, where $d_{k,b}$ (in meter) is the distance between MS_k and BS_b. We assume that the receive SNR of the cell-edge MS is 10 dB. The small scale fading channels between BSs and MSs are i.i.d. Rayleigh channels. The codebooks used for quantizing the per-cell channels are obtained by random vector quantization (RVQ). The codebook size for feeding back each per-cell channels is set as four bits. All simulation results are obtained by averaging over 1000 realizations of the small scale fading channels. We consider that two MSs are activated in each sector and the three BSs cooperatively serve the six MSs simultaneously.

To clearly observe the impact of large scale fading gains on the performance of different codeword selection methods, we first consider a special scenario with the MS locations shown in Fig. 1. Specifically, the two MSs in the same sector are located in the same place and the MS-groups in different sectors are at the same distance from their local BSs, which is denoted by d_1. In this way, we only need to show the performance of one MS. The performance under practical random MS locations will be shown later in Fig. 4.

5.1.2 Performance Comparison of Different Codeword Selection Methods

To show the impact of different criteria for codeword selection on the performance, we first provide the results with exhaustive searching. Note that the codewords cannot be selected to maximize the actual data rate during downlink transmission in practical systems, due to the fundamental challenge of FDD systems

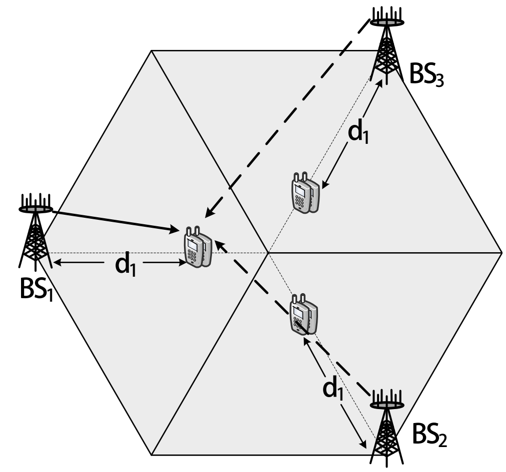

Figure 1: An example of CoMP system, where the solid line denotes local channel while the dash lines denote cross channels for an MS. The cell radius is 250 m. The MSs in the same cell are co-located in the same place, and the user-groups in different cells are at the same distance from their local BSs.

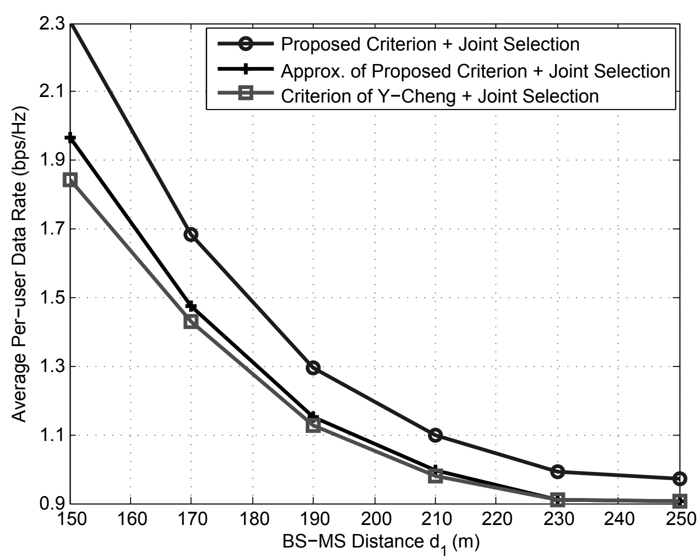

Figure 2: Average per-user data rate versus the BS-MS distance d_1. Each user is equipped with two antennas and two data streams are transmitted to each user, i.e., $N_R = 2$, $d_k = 2$.

that the users do not have the CSI of other users. Although we can simulate the performance of the codeword selection by maximizing the actual data rate, which can serve as an upper bound for comparison, the codeword selection to maximize actual downlink data rate requires an exhaustive searching over the N_B per-cell codebooks of all K users, whose complexity is on the order of $\mathcal{O}(\prod_{k=1}^{K}\prod_{b=1}^{N_B}2^{B_{k,b}})$. Under the considered simulation settings, the order of complexity is as high as $\mathcal{O}(2^{72})$, which cannot be afforded in simulation. Moreover, such an upper bound is far from achievable in practice; therefore, we do not provide its simulation results.

In Fig. 2, the average per-user data rates versus BS-MS distance d_1 under three codeword selection criteria are compared. In the simulation, each MS is equipped with two antennas and two data streams are transmitted to each MS, i.e., $N_R = 2$, $d_k = 2$. The codewords are selected by exhaustive searching according to the following three criteria: 1) the proposed criterion in (12), with the legend "Proposed Criterion + Joint Selection"; 2) minimizing the approximated objective function in (16), with the legend "Approx. of Proposed Criterion + Joint Selection"; and 3) the criterion considered in [3], which is shown in (19), with the legend "Criterion of Y-Cheng + Joint Selection". We can observe that the per-user data rate achieved by our proposed criterion is the highest. Maximizing the approximation of the proposed criterion causes performance loss, but it still outperforms the criterion proposed in [3]. The performance gain of the proposed criterion over the criterion in [3] increases when the MSs move from cell edge to cell center, i.e., the value of d_1 at x-axis decreases. This is because the codeword selection criterion in [3] does not exploit the large scale fading gains of CoMP channel during codeword selection.

To evaluate the performance of the proposed serial codeword selection method, in Fig. 3 the average per-user data rates versus BS-MS distance d_1 under three different codeword selection methods are compared. In the simulation, each MS is equipped with a single antenna, i.e., $N_R = 1$, $d_k = 1$. The three codeword selection methods are: 1) optimal selecting codeword method by exhaustive searching according to the proposed criterion in (12), with the legend "Proposed Criterion + Joint Selection"; 2) the proposed serial codeword selection, with the legend "Proposed Criterion + Serial Selection"; and 3) the low-complexity method proposed in [3] with different complexities, with the legend "Criterion and Selection Method of Y-Cheng". As expected, the per-user data rate achieved with the optimal codeword selection method is the highest, while the good performance is paid by high order of complexity as $\mathcal{O}(4096)$. The performance of serial codeword selection method is close to the joint codeword selection, and the performance gap decreases when the MSs move from cell edge to cell center. Despite such a small performance loss, the complexity has been dramatically reduced, whose order is $\mathcal{O}(48)$ and is about 1/85 of the optimal codeword selection method. As for the low-complexity method proposed in [3], when its complexity is set the same as the serial codeword selection, the method reduces to selecting each per-cell codeword that has the minimal chordal distance with the per-cell channel vector, and can be regarded as an independent codeword selection for the per-cell channels. The independent per-cell codeword selection method performs the worst, since it ignores the inter-cell phase information during the selection. When the codeword selection complexity of method in [3] is increased to two times larger than the serial codeword selection, the performance is substantially improved but is still inferior to the serial codeword selection.

To evaluate the performance of the proposed criterion and the proposed serial codeword selection method in a more realistic user distribution, in Fig. 4 we provide the average per-user data rate when six MSs are randomly distributed in a 10 dB "cell-edge region", where $\min\limits_{l\neq b_k}\frac{\alpha_{k,b_k}^2}{\alpha_{k,l}^2}$ for MS_k is less than 10 dB. This corresponds to randomly scheduling the MSs for transmission. In practice, any well-designed scheduler will perform better than a random scheduler. Four codeword selection methods are compared, which are: 1) exhaustively searching codewords according to (12), 2) exhaustively searching codewords according to (19), i.e., the criterion of Y-Cheng; 3) selecting codewords by our proposed serial codeword selection, and 4) the low complexity codeword selection method in [3]. The legends are the same as before. For a fair comparison, the complexity of the method in [3] is set the same as that of the serial codeword selection method. We can see that the serial codeword selection method yields approximately the optimal result with exhaustive searching but with quite low complexity, and outperforms the method in [3] with the same complexity.

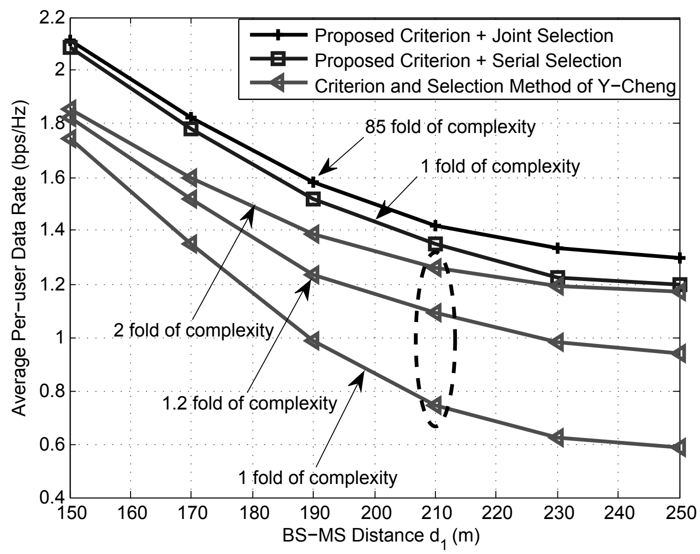

Figure 3: Average per-user data rate versus the BS-MS distance d_1. Each user is equipped with single antenna and single data stream is transmitted to each user, i.e., $N_R = 1$, $d_k = 1$.

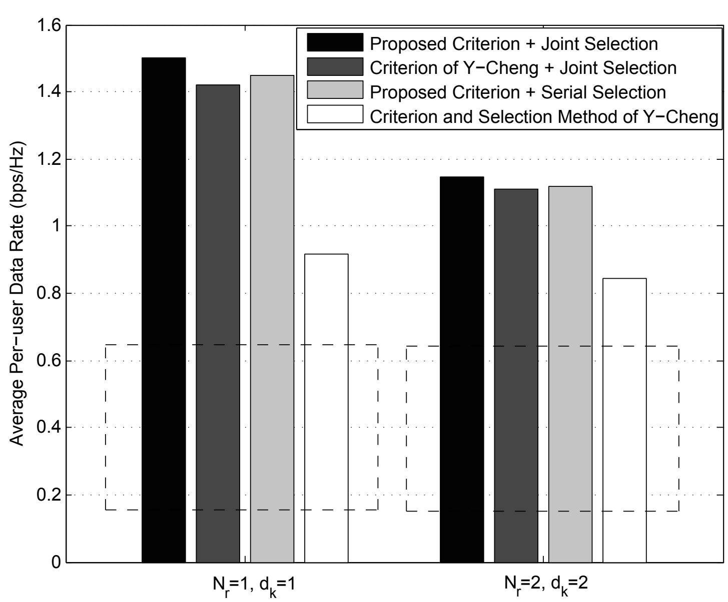

Figure 4: Average per-user data rate under different configurations.

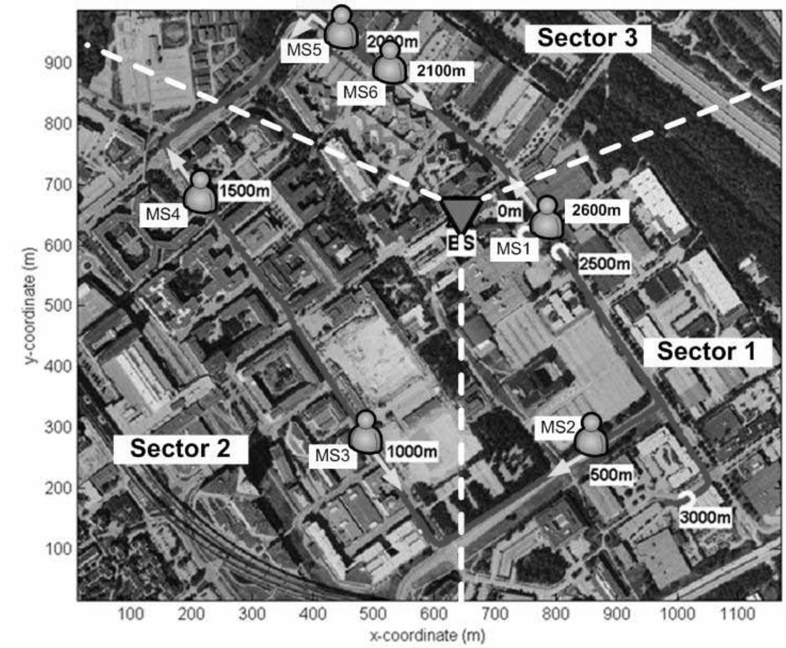

Figure 5: Scenario and driving route in the multi-sector measurement campaign. The six MSs are initially located at locations labeled as 500 m, 1000 m, 1500 m, 2000 m, 2100 m and 2600 m.

5.2 Performance Comparison in a Measured Urban Environment

Finally, we evaluate the performance of different codeword selection methods in a realistic multi-sector scenario based on channel measurements. The measurement was performed in an urban macrocellular environment at Kista, near Stockholm, using one four-antenna MS and a three-sector BS site, where each sector is equipped with a cross-polarized antenna pair. The equipments and setup are described in detail in [10], which are omitted here for brevity. Although the MS in the measurement is equipped with four antennas, here we only consider one receive antenna, in order to support multiple MSs in each sector and apply the method in [3]. The collected channel measurement is used to generate our evaluation scenario, where six MSs are moving around in the area covered by the three BS sectors, as shown in Fig. 5. To study the impact of codeword selection methods on the performance of individual MSs, the six MSs are initially placed at the positions shown in Fig. 5 and moved at a constant speed (of approximately 30 km/hr) according to the indicated directions. The transmit power of each BS and the thermal noise power of MSs are set as 46 dBm and -96 dBm, which are in accordance with the LTE specification [18]. Each data rate sample is obtained by averaging over 162 frequency samples and 50 time domain channel samples.

The performance is compared in Fig. 6, where the data rates of the six MSs versus their moving distances are provided. The four codeword selection methods are the same as that considered in Fig. 4. From the results we can observe that when exhaustively searching the codewords, the proposed criterion in (12) always outperforms the criterion proposed in [3]. The performance gap differs for various MSs and different locations of each MS. The performance of serial codeword selection almost overlaps with that from the optimal codeword selection method by exhaustively searching according to (12), and is superior to the low complexity method in [3], no matter where the MS is located. This results further substantiate the good performance of the proposed low-complexity method.

6 Conclusions

In this paper, we studied codeword selection for limited feedback CoMP-JP systems with per-cell codebook. A unified codeword selection criterion was provided for an arbitrary number of antennas and an arbitrary number of data streams, which degenerates to various selection criteria under different configurations, and outperforms other criterion for CoMP known in literature. By exploiting the imbalance of average channel gains from multiple BSs to an MS, we proposed a low-complexity codeword selection method. The proposed codeword selection criterion and method were evaluated in a measured urban environment and through simulations. The results showed that the serial codeword selection method performs closely to the optimal codeword selection that maximizes the estimated data rate with exhaustive searching, and outperforms existing scheme with the same complexity.

Acknowledgements

The authors thank Dr. Jonas Medbo of Ericsson AB, Stockholm, for providing the measurement data used in this work.

References

1 Karakayali M, Foschini G, Valenzuela R. Network coordination for spectrally efficient communications in cellular systems. IEEE Wirel Commun Mag, 2006, 13: 56–61
2 Gesbert D, Hanly S, Huang H, et al. Multi-cell MIMO cooperative networks: A new look at interference. IEEE J Sel Areas Commun, 2010, 28: 1380–1408
3 Cheng Y, Lau V K N, Long Y. A scalable limited feedback design for network MIMO using per-cell codebook. IEEE Trans Wirel Commun, 2010, 9: 3093–3099
4 Love D, Heath R W Jr, Lau V K N, et al. An overview of limited feedback in wireless communication systems. IEEE J Sel Areas Commun, 2008, 26: 1341–1365
5 Jindal N. MIMO broadcast channels with finite rate feedback. IEEE T Inform Theory, 2006, 52: 5045–5059
6 Ravindran N, Jindal N. Limited feedback-based block diagonalization for the MIMO broadcast channel. IEEE J Sel Areas Commun, 2008, 26: 1473–1482
7 Boccardi F, Huang H, Trivellato M. Multiuser eigenmmode transmission for MIMO broadcast channels with limited feedback. Proceedings of IEEE International Workshop on Signal Processing Advances for Wireless Communications, Helsinki, Finland, 2011. 1–5
8 Jindal N. Antenna combining for the MIMO downlink channel. IEEE Trans Wirel Commun, 2008, 7: 3834–3844
9 Trivellato M, Boccardi F, Huang H. On transceiver design and channel quantization for downlink multiuser MIMO systems with limited feedback. IEEE J Sel Areas Commun, 2008, 26: 1494–1504
10 Tian R, Wu B, Lau B K, et al. On MIMO performance enhancement with multisector cooperation in a measured urban environment. IEEE Antenn Wirel PR, 2012, 11: 488–491
11 Tse D, Viswanath P. Fundamentals of wireless communication. Cambridge University Press, 2004
12 Spencer Q, Swindlehurst A L, Haardt M. Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels. IEEE T Signal Proces, 2004, 52: 461–471
13 Zhang J, Chen R, Andrews J, et al. Networked MIMO with clustered linear precoding. IEEE Trans Wirel Commun, 2009, 8: 1910–1921
14 Zhang R. Cooperative multi-cell block diagonalization with per-base-station power constraints. IEEE J Sel Areas Commun, 2010, 28: 1435–1445
15 Bayesteh A, Khandani A K. On the user selection for MIMO broadcast channels. IEEE T Inform Theory, 2008, 56: 1086–1107
16 Love D, Heath R W Jr, Strohmer T. Grassmannian beamforming for multiple-input multiple-output wireless systems. IEEE T Inform Theory, 2003, 49: 2735–2747
17 Dai W, Dai Y, Rider B. Quantization bounds on grassmann manifolds and applications to MIMO communications. IEEE T Inform Theory, 2008, 54: 1108–1123
18 3GPP Long Term Evolution (LTE). Further advancements for E-UTRA physical layer aspects. TSG RAN TR 36.814, 2009

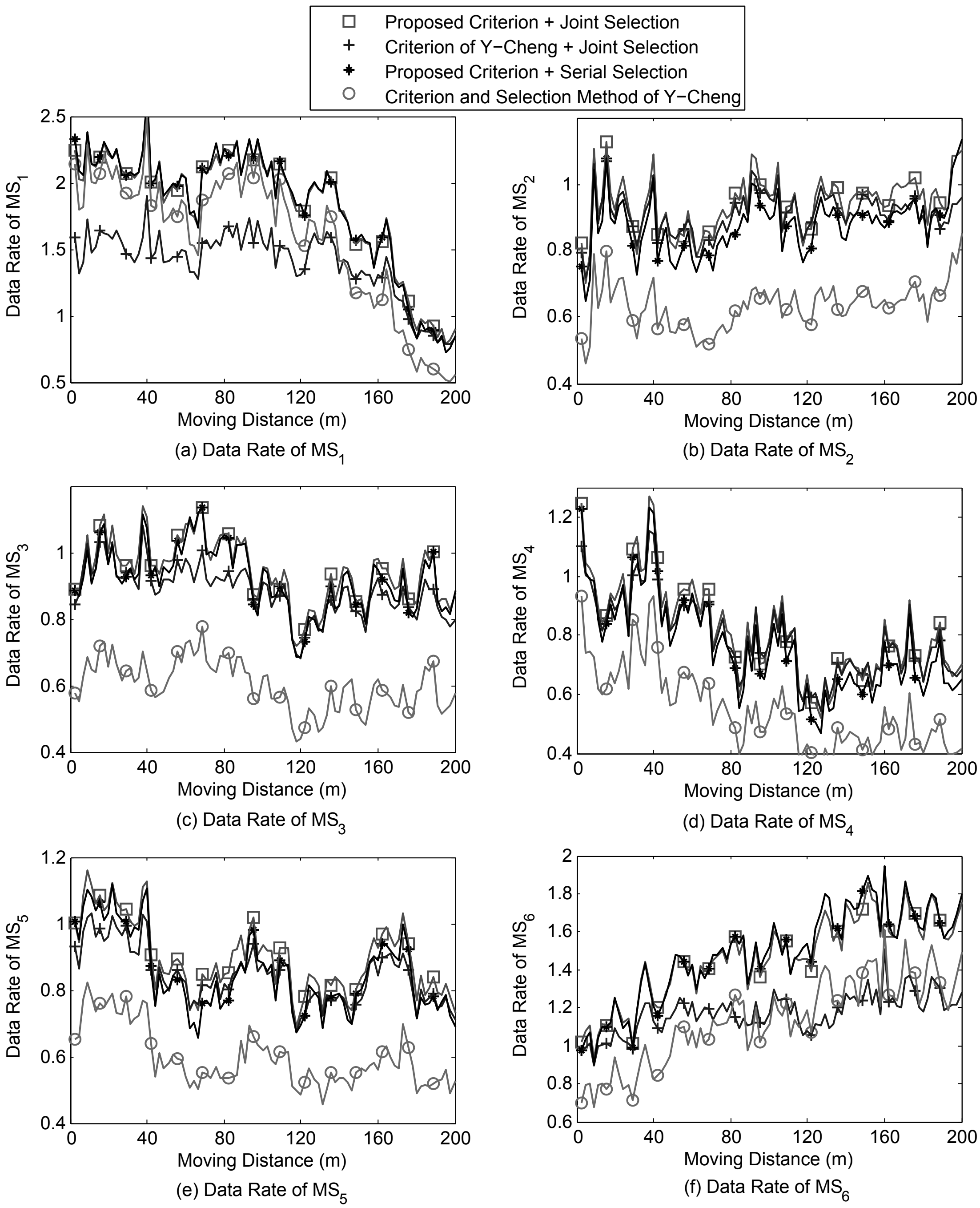

Figure 6: The data rates of six MSs versus their moving distances.