

Article

Hybrid MU-MIMO Precoding Based on K-Means User Clustering

Razvan-Florentin Trifan ^{1,2}, Andrei-Alexandru Enescu ¹ and Constantin Paleologu ^{1,*} 

¹ Department of Telecommunications, University POLITEHNICA of Bucharest, 1-3, Iuliu Maniu Blvd., 061071 Bucharest, Romania

² InfoVista SAS, 23 Carnot Av., 91300 Massy, France

* Correspondence: pale@comm.pub.ro; Tel.: +40-21-402-4634

Received: 31 May 2019; Accepted: 19 July 2019; Published: 23 July 2019



Abstract: Multi-User (MU) Multiple-Input-Multiple-Output (MIMO) systems have been extensively investigated over the last few years from both theoretical and practical perspectives. The low complexity Linear Precoding (LP) schemes for MU-MIMO are already deployed in Long-Term Evolution (LTE) networks; however, they do not work well for users with strongly-correlated channels. Alternatives to those schemes, like Non-Linear Precoding (NLP), and hybrid precoding schemes were proposed in the standardization phase for the Third-Generation Partnership Project (3GPP) 5G New Radio (NR). NLP schemes have better performance, but their complexity is prohibitively high. Hybrid schemes, which combine LP schemes to serve users with separable channels and NLP schemes for users with strongly-correlated channels, can help reduce the computational burden, while limiting the performance degradation. Finding the optimum set of users that can be co-scheduled through LP schemes could require an exhaustive search and, thus, may not be affordable for practical systems. The purpose of this paper is to present a new semi-orthogonal user selection algorithm based on the statistical K-means clustering and to assess its performance in MU-MIMO systems employing hybrid precoding schemes.

Keywords: 5G NR (New Radio); Non-linear Precoding (NLP); hybrid precoding; Tomlinson–Harashima Precoding (THP); Regularized Zero Forcing (RZF); Block Diagonalization (BD); gaussian angular correlation; spatial compatibility; MU-MIMO; K-means clustering; user selection

1. Introduction

Multi-User (MU) Multiple-Input-Multiple-Output (MIMO) increases network capacity by multiplexing spatially-separated users in the same frequency and time resources. It requires transmission precoding, as a random geographic location of the receivers does not enable joint decoding. Thus, it is challenging to send independent data streams simultaneously to a set of properly-selected users to attain the spatial multiplexing gains offered by MU-MIMO. A recent evolution of MU-MIMO is known as massive MIMO or large-scale MIMO [1] and is characterized by Base Stations (BSs) with hundreds of antennas sending different data streams to tens of users. Massive MIMO has been identified as the air interface technology capable of addressing the massive capacity requirement of 5G networks.

Linear Precoding (LP) schemes for MU-MIMO eliminate the inter-user interference by projecting the intended user's signal into the null space of the other users [2]. When their channels are strongly correlated, i.e., for closely-separated users, the projection operation makes it almost impossible to separate the received signals and results in high capacity loss [3]. Existing MU-MIMO schedulers estimate the spatial separation between users and try to serve the ones with near-orthogonal channels. The optimum set of users can be found through an exhaustive search. The combinatorial nature of

these problems makes them computationally infeasible due to an expected large number of users and antennas in massive MIMO systems. Various sub-optimal user grouping strategies, like null space projection or spatial clustering classification, have already been evaluated in the literature, including [2] and the references therein.

To reduce the channel estimation overhead in massive MIMO Frequency Division Duplex (FDD) systems, a user grouping algorithm based on K-means clustering was proposed in [4,5]. The performance of a two-stage precoding algorithm with various distance metrics for user grouping was tested; however, all require complex matrix computations. The Block Diagonalization (BD) linear precoder [6,7] is used to cancel the inter-group interference. The results are not accurate enough, as the number of clusters is assumed to be known prior, and there are no insights of the spatial separation between the co-scheduled groups. In addition, the subspaces of groups (created using the average covariance matrix of the users in a group) are slightly different from the channel eigenspace of users. As such, the inter-group interference is not entirely canceled. Lastly, the employed intra-group precoder is based on linear regularized zero forcing [8]. While linear precoding mechanisms have a relatively low implementation complexity, they may not exhibit a good performance for highly-correlated channels, which is the case of users inside the same group [3,9]. Non-Linear Precoding (NLP) schemes, such as Tomlinson–Harashima Precoding (THP) [10], proposed in the standardization phase for 5G NR (New Radio) [11,12], or hybrid precoding [13], could behave better in such scenarios.

A user grouping algorithm based on an optimized K-means clustering is presented in the paper. This further reduces the complexity of the method proposed in [4,5] and ensures that, at a given time, at least one user in one group is scheduled. In contrast to the related work, the number of co-scheduled clusters is dynamically chosen, such that users of different clusters are separated enough in space to be served through an LP scheme. As users inside the same group have highly correlated channels, they will be scheduled in the same time-frequency resource through a second-stage non-linear precoder.

To assess the particularities of the proposed algorithm, the paper is organized into the following sections. Section 2 describes the basic principles of precoding, by identifying three types of precoding schemes. Section 3 defines a model for the spatial correlation in MU-MIMO channels and a metric for evaluating the separation between channels. The proposed user grouping solution is analyzed in detail in Section 4. Section 5 presents the simulation's parameters used to test the performance of the proposed algorithm. The clustering of users and throughput performance of the MU-MIMO system are then evaluated in Section 6 for various user distributions and required angle separations. Finally, Section 7 presents the conclusion and research directions for future works.

2. Precoding Schemes

In this section, the precoding schemes are grouped into three categories, i.e., linear, non-linear, and hybrid.

2.1. Linear Precoding

Linear precoding schemes, as used in Long-Term Evolution (LTE), suppress the inter-user interference through channel inversion. Techniques like Zero Forcing (ZF) and Regularized Zero Forcing (RZF) for single-antenna User Equipment (UEs) or Block Diagonalization (BD) for multiple-antenna UE have reduced complexity [7].

The channel matrix with complex coefficients corresponding to N_T transmit antennas and K single-antenna users is described by $\mathbf{H} \in \mathbb{C}^{K \times N_T}$.

The precoder identified by the matrix $\mathbf{W} \in \mathbb{C}^{N_T \times K}$ and represented in Figure 1 can be defined by the following equations:

$$\mathbf{W}_{ZF} = \beta \mathbf{H}^{-1}, \quad (1)$$

where $\mathbf{x} \in \mathbb{C}^{K \times 1}$ is the set of complex symbols intended for every K user, $\mathbf{y} \in \mathbb{C}^{K \times 1}$ is the received array of symbols by the K users, and $\mathbf{n} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma_n^2 \mathbf{I}_K)$ is the vector of Additive White Gaussian Noise (AWGN) of variance σ_n^2 . It follows a circularly-symmetric complex Gaussian distribution with zero mean and correlation matrix $\sigma_n^2 \mathbf{I}_K$.

$$\begin{aligned} \mathbf{W}_{RZF} &= \beta \times \arg \min_{\mathbf{W}} \mathbb{E} \{ \|\beta^{-1}(\mathbf{H}\mathbf{W}\mathbf{x} + \mathbf{n}) - \mathbf{x}\|^2 \} \\ &= \beta \mathbf{H}^H \left(\mathbf{H}\mathbf{H}^H + \frac{N_T \sigma_n^2}{P_{Tx}} \mathbf{I} \right)^{-1}, \end{aligned} \tag{2}$$

where $P_{Tx} = \mathbb{E} \{ \mathbf{x}^H \mathbf{x} \}$ is the total transmit power, $\mathbb{E} \{ x \}$ is the expectation of a random variable x , $\|\mathbf{x}\|$ is the L_2 -norm of a vector \mathbf{x} , superscript H denotes the transpose complex conjugate, and β is a normalization factor, such that the total transmit power does not change after precoding:

$$\beta = \sqrt{\frac{N_T}{\text{tr}(\mathbf{W}\mathbf{W}^H)}}, \tag{3}$$

where $\text{tr}(\mathbf{X})$ is the trace of a square matrix \mathbf{X} .

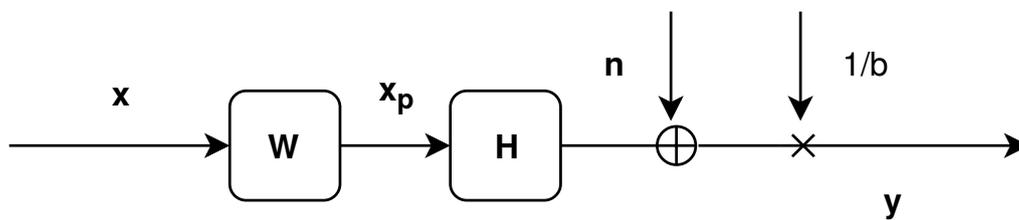


Figure 1. Linear precoding scheme [3].

The previously-mentioned schemes are used for users with a single antenna. The signal transmitted for other users is treated as interference, and the precoder attempts to cancel it.

When users are equipped with multiple antennas, similar methods can be used. In this case, the signal intended for the same user, but for different antennas, is also treated as interference. As there is much more interference to cancel, noise is amplified at the receiver side.

In these scenarios, the BD precoder performs better. It works by decomposing the MU-MIMO downlink channel into multiple parallel independent single-user MIMO downlink channels. The precoding matrix is determined such that the intended signal is projected in the null space of all other users' channel matrices. The remaining interference after the inter-user precoding can be canceled with any decoding or precoding techniques.

The received signal by K users equipped with N_R antennas is:

$$\begin{aligned} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \dots \\ \mathbf{y}_K \end{bmatrix} &= \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_1 & \dots & \mathbf{H}_1 \\ \mathbf{H}_2 & \mathbf{H}_2 & \dots & \mathbf{H}_2 \\ \dots & \dots & \dots & \dots \\ \mathbf{H}_K & \mathbf{H}_K & \dots & \mathbf{H}_K \end{bmatrix} \begin{bmatrix} \mathbf{W}_1 \mathbf{x}_1 \\ \mathbf{W}_2 \mathbf{x}_2 \\ \dots \\ \mathbf{W}_K \mathbf{x}_K \end{bmatrix} + \begin{bmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \\ \dots \\ \mathbf{n}_K \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{H}_1 \mathbf{W}_1 & \mathbf{H}_1 \mathbf{W}_2 & \dots & \mathbf{H}_1 \mathbf{W}_K \\ \mathbf{H}_2 \mathbf{W}_1 & \mathbf{H}_2 \mathbf{W}_2 & \dots & \mathbf{H}_2 \mathbf{W}_K \\ \dots & \dots & \dots & \dots \\ \mathbf{H}_K \mathbf{W}_1 & \mathbf{H}_K \mathbf{W}_2 & \dots & \mathbf{H}_K \mathbf{W}_K \end{bmatrix} + \begin{bmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \\ \dots \\ \mathbf{n}_K \end{bmatrix}, \end{aligned} \tag{4}$$

where $\mathbf{H}_u \mathbf{W}_{k,u \neq k}$ is the effective channel corresponding to the signal sent to user k and received by user u . This interferes with user k when $\mathbf{H}_u \mathbf{W}_k \neq \mathbf{0}_{N_R \times N_R}$. The effective channel matrix needs to be diagonalized in order to cancel the inter-user interference.

$$\mathbf{H}_u \mathbf{W}_k = \mathbf{0}_{N_R \times N_R}, \quad \forall u \neq k. \tag{5}$$

To prevent any power increase after precoding, the BD precoder matrix has to be unitary.

Assuming a matrix containing the channel gains of all users except u ,

$$\tilde{\mathbf{H}}_u = [\mathbf{H}_1, \dots, \mathbf{H}_{u-1}, \mathbf{H}_{u+1}, \dots, \mathbf{H}_K]^T, \tag{6}$$

when $N_T = K \times N_R$, the Singular-Value Decomposition (SVD) of $\tilde{\mathbf{H}}_u$ is:

$$\tilde{\mathbf{H}}_u = \tilde{\mathbf{U}}_u \tilde{\mathbf{\Lambda}}_u [\tilde{\mathbf{V}}_u^{\text{non-zero}} \tilde{\mathbf{V}}_u^{\text{zero}}]^H, \tag{7}$$

where $\tilde{\mathbf{V}}_u^{\text{non-zero}} \in \mathcal{C}^{(K \times N_R - N_R) \times N_T}$ and $\tilde{\mathbf{V}}_u^{\text{zero}} \in \mathcal{C}^{N_R \times N_T}$ contain singular vectors corresponding to the non-zero and zero singular values, respectively.

$$\begin{aligned} \tilde{\mathbf{H}}_u \tilde{\mathbf{V}}_u^{\text{zero}} &= \tilde{\mathbf{U}}_u [\tilde{\mathbf{\Lambda}}_u^{\text{non-zero}} \mathbf{0}] \begin{bmatrix} (\tilde{\mathbf{V}}_u^{\text{non-zero}})^H \\ (\tilde{\mathbf{V}}_u^{\text{zero}})^H \end{bmatrix} \tilde{\mathbf{V}}_u^{\text{zero}} \\ &= \tilde{\mathbf{U}}_u \tilde{\mathbf{\Lambda}}_u^{\text{zero}} (\tilde{\mathbf{V}}_u^{\text{non-zero}})^H \tilde{\mathbf{V}}_u^{\text{zero}} \\ &= \tilde{\mathbf{U}}_u \tilde{\mathbf{\Lambda}}_u^{\text{zero}} \mathbf{0} \\ &= \mathbf{0}. \end{aligned} \tag{8}$$

As $\tilde{\mathbf{V}}_u^{\text{zero}}$ lies in the null space of $\tilde{\mathbf{H}}_u$, the signal transmitted in the direction of $\tilde{\mathbf{V}}_u^{\text{zero}}$ is only received by user u . Its corresponding precoder is $\mathbf{W}_u = \tilde{\mathbf{V}}_u^{\text{zero}}$.

2.2. Non-Linear Precoding Schemes

Linear precoding has relatively low complexity; however, it has been shown that the sum rate of channel inversion is limited when increasing the number of users. Furthermore, linear precoding does not perform well in scenarios with high UE density and correlated channels.

Non-linear precoding schemes are valid alternatives, which instead of spatially separating the users, they eliminate the interference through non-linear operations. The non-linear concept was introduced in Dirty Paper Coding (DPC) [14]. It can achieve the maximum sum rate of the system when the full Channel State Information (CSI) is available at the transmitter side. The DPC's pre-subtraction of the non-causally known interference has a high complexity, and as a result, sub-optimum non-linear precoding schemes such as Tomlinson–Harashima Precoding (THP) and Vector Perturbation (VP) were proposed as alternatives with lower complexity [10].

2.2.1. THP Precoding

The concept of THP is derived from the Decision Feedback Equalizer (DFE) technique. There, the interference resulting from the previously-estimated received symbols was reconstructed and subtracted from the actual received symbol. THP is represented in Figure 2. The power growth resulting from the successive cancellation step is limited by using the modulo function.

$$\text{mod}(x) = x - 2\sqrt{M} \left\lfloor \frac{s + \sqrt{M} + j\sqrt{M}}{2\sqrt{M}} \right\rfloor, \tag{9}$$

where:

$$\mathbf{x}_p = [x_{pk}] = \text{mod} \left(x_k - \sum_{l=1}^{K-1} b_{kl} x_{pl} \right), \quad \forall k \in 1, \dots, K \tag{10}$$

and the feedback matrix $\mathbf{B} = [b_{kl}]_{k,l=1,2,\dots,K}$ is a unitary lower diagonal matrix, obtained through a QR decomposition of \mathbf{H} , such that:

$$\mathbf{B} = \begin{bmatrix} 1/l_{11} & 0 & \cdots & 0 \\ 0 & 1/l_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1/l_{KK} \end{bmatrix} \begin{bmatrix} l_{11} & 0 & \cdots & 0 \\ l_{21} & l_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ l_{K1} & l_{K2} & \cdots & l_{KK} \end{bmatrix} \tag{11}$$

and $\mathbf{x}_{THP} = \mathbf{Q}^H \mathbf{x}$.

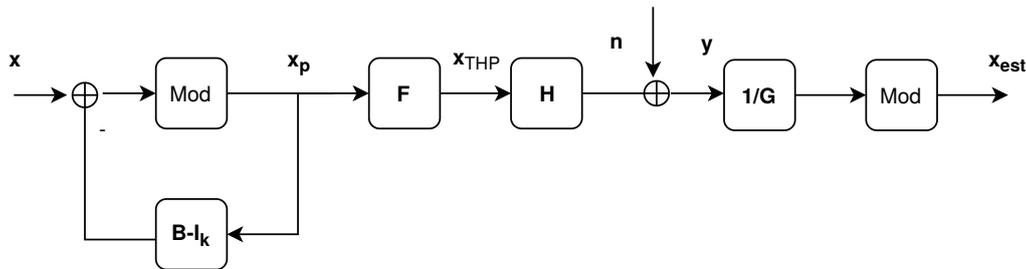


Figure 2. THP block scheme [3].

At the receiver, the user’s k signal is firstly divided by $g_k = l_{kk}$, then passed through the modulo operation to be then sent to the Quadrature Amplitude Modulation (QAM) demodulator.

2.2.2. Vector Perturbation Precoding

Vector perturbation works similarly to THP, as the transmitted symbols are “perturbed” and moved from the original to a new constellation of points. On the receiver side, a diagonal matrix is used to scale the received signal, which is then reconstructed through a modulo operation.

2.3. Hybrid Precoding

Hybrid precoding, a mix of linear and non-linear precoding schemes, was proposed in [11–13] to enhance the robustness of non-linear precoding against channel imperfections and to reduce the complexity required for large-scale antenna systems.

Considering K users divided into G groups, the precoding mechanism is implemented in two stages. The first stage consists of a linear precoder used to eliminate the inter-group interference and assumes that the users from different groups have sufficiently uncorrelated channels. This stage can be realized through BD using long-term CSI measurements. Based on the degree of correlation between the users of the same group, linear or non-linear precoding schemes can be employed inside every group, in the second stage.

Figure 3 depicts a hybrid precoding scheme that uses BD and THP:

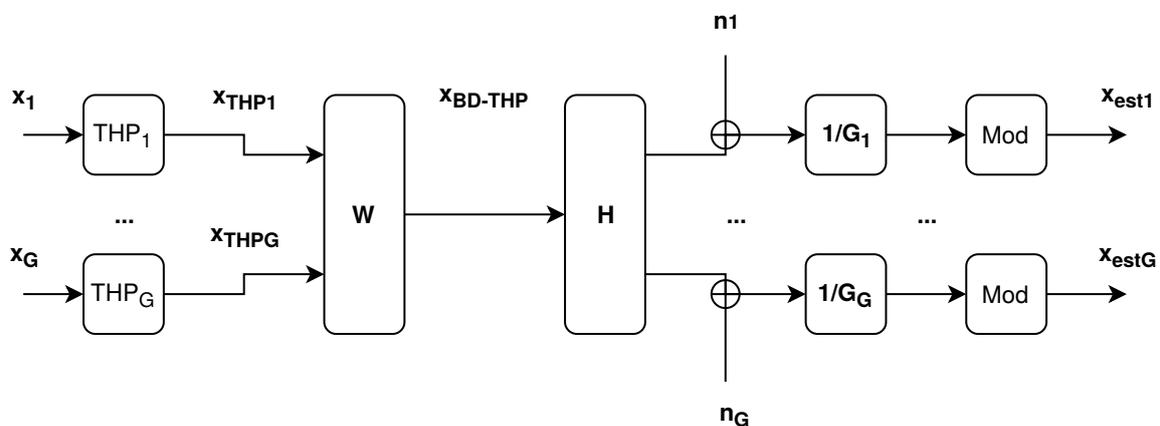


Figure 3. Hybrid precoder block scheme [3].

2.4. Complexity Evaluation

The complexity can be evaluated by counting the number of Floating Point Operations (FLOPs) required for multiplications and additions of complex values. A multiplication and an addition requires six and two FLOPs, respectively. The computational complexity of the SVD decompositions required by the BD-THP precoder is neglected, being computed less when relying on statistical CSI. The complexity of the QR decomposition of channel matrix \mathbf{H} is $8N_T K^2 - 8K^3/3$. Generating the matrix $\mathbf{H}\mathbf{H}^H$ for the RZF precoder uses $K(K+1)(4N_T - 1)$ FLOPs due to the Hermitian property of the Gram matrix. More details on how to evaluate the complexity calculations can be found in [13].

The complexity of various precoding schemes is represented in Table 1 [3,13], for $K_g = K/G$ users in each group.

Table 1. Precoder’s complexity [3,13].

Precoder	Number of FLOPs
RZF	$4K^3 + 2KN_T(4K - 1) + K(4N_T - 1)(K + 1) + 2N_T(4K - 1) + 8K^2 + 7K$
THP	$16K^3/3 + 2K(4KN_T - K + 2) + 2T[2K + 2K^2 + N_T(4K - 1) - 4]$
BD-THP	$40GK_g^3/3 - 4GK_g^2 + 2GK_g - 2GK_g N_T + 16GK_g^2 N_T + T(4GK_g^2 + 4GK_g + 8K_g N_T - 8G - 2N_T)$

In Figure 4, the equations above are plotted for $N_T = 16$ transmit antennas and $K = 16$ single-antenna users grouped into $G = 4$ groups. Although the BD-THP scheme is a good tradeoff between the THP and the RZF schemes in terms of the sum rate (as will be seen in the following sections), its complexity is below both of the previous schemes. The computation complexity required to generate T precoded data vectors can change the ranking of schemes (for low values of T); however, the BD-THP precoder will always show the lowest complexity among them.

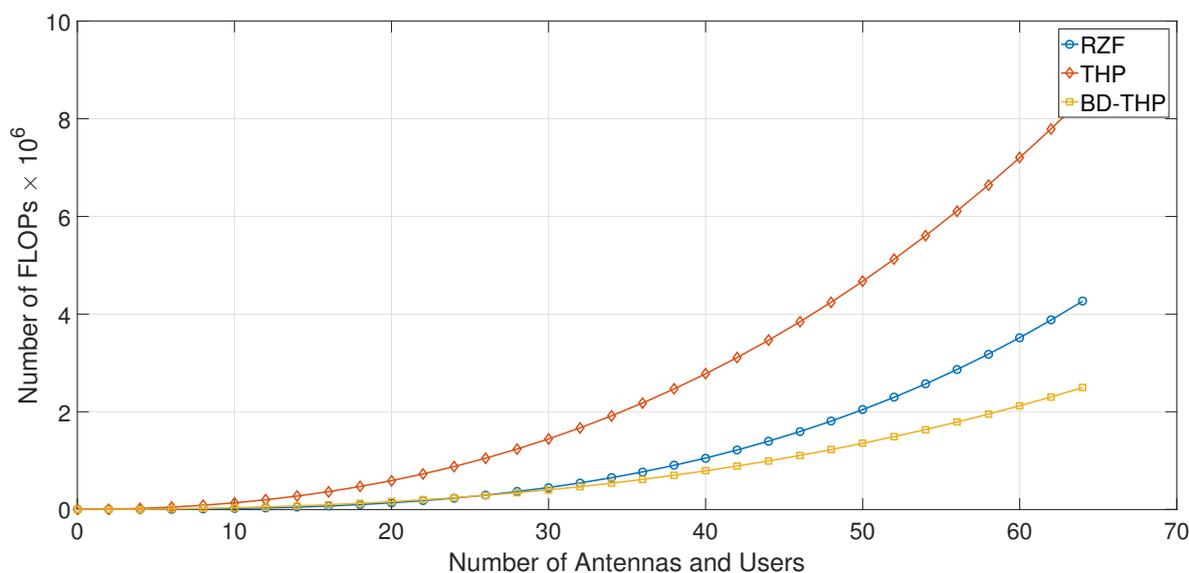


Figure 4. Precoders complexity versus the number of antennas and users, for $N_T = K$.

3. Spatial Compatibility and Channel Modeling

A set of users is spatially compatible if their channels, \mathbf{h}_k , can be separated in space through precoding. The purpose of this section is to define a metric that can evaluate how spatially compatible

two users are. Assuming a spatial correlation model at the transmission, it can be shown that the separation between the channels of two users will depend only on their spatial correlation matrices and, in turn, on the angle between them relative to the antenna system.

3.1. Channels Separation Metric

The spatial channel correlation is important for multiple-antenna users (Single User (SU)-MIMO communications); however, in the context of single-antenna users, it is the set of spatial correlation matrices of users that determine the network performances.

A metric used to quantify how separated the channels of two users are was described in [15,16]. Assuming that \mathbf{h}_i and \mathbf{h}_k are the channels of two users, the channel separation metric is defined as:

$$\cos(\angle(\mathbf{h}_i, \mathbf{h}_k)) \triangleq \mathbb{V} \left\{ \frac{\mathbf{h}_i^H \mathbf{h}_k}{\sqrt{\mathbb{E}\{|\mathbf{h}_i|^2\} \mathbb{E}\{|\mathbf{h}_k|^2\}}} \right\}, \quad (12)$$

where $\cos(\angle(\mathbf{h}_i, \mathbf{h}_k)) \in [0, 1]$ and $\mathbb{V}\{x\}$ is the variance of a random variable x .

The metric indicates how efficiently the transmitter can serve user i without affecting user k , and vice versa. Strongly-correlated channels exhibit higher values, whereas for Rayleigh uncorrelated fading, the variance is $1/N_T$ and decreases when the number of antennas increases. In general, the variance depends on the spatial channel correlation and equals zero when the users have orthogonal correlation matrices, $\angle(\mathbf{h}_i, \mathbf{h}_k) = \frac{\pi}{2}$.

In MU-MIMO systems, a set of users is called ϵ -orthogonal, if $\cos(\angle(\mathbf{h}_i, \mathbf{h}_k)) < \epsilon$ for every $i \neq k$. The users can be grouped into disjoint sets according to a desired threshold ϵ , such that semi-orthogonal sets could be scheduled over independent resources.

3.2. Correlation Model

The authors in [17] mentioned that the channel separation metric should also be a function of the spatial correlation between \mathbf{h}_k and \mathbf{h}_i and considered the inner correlation of each multi-antenna user. However, the precoding performance was primarily affected by the correlation between transmit antennas, whereas receive antenna correlation had an insignificant impact on the precoding design [18]. In a MIMO context, a spatially-correlated random channel vector \mathbf{h}_k can be created using Karhunen–Loeve expansion [15]:

$$\mathbf{h}_k = \mathbf{R}_k \hat{\mathbf{e}}_k = \mathbf{U}_k \mathbf{D}_k^{\frac{1}{2}} \mathbf{U}_k^H \hat{\mathbf{e}}_k \sim \mathbf{U}_k \mathbf{D}_k^{\frac{1}{2}} \mathbf{e}_k, \quad (13)$$

where $\hat{\mathbf{e}}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_{N_T}, \mathbf{I}_{N_T})$ and $\mathbf{e}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_r, \mathbf{I}_r)$. The eigenvalue decomposition of $\mathbf{R}_k \in \mathbb{C}^{N_T \times N_T}$ is $\mathbf{R}_k = \mathbf{U}_k \mathbf{D}_k \mathbf{U}_k^H$, where $\mathbf{D}_k \in \mathbb{R}^{r \times r}$ is a diagonal matrix containing $r = \text{rank}(\mathbf{R}_k)$ positive non-zero eigenvalues of \mathbf{R}_k and $\mathbf{U}_k \in \mathbb{C}^{N_T \times r}$ contains the associated eigenvectors. The last part means that the distributions of \mathbf{h}_k of $\mathbf{U}_k \mathbf{D}_k^{\frac{1}{2}} \mathbf{e}_k$ are identical.

It can be shown that the spatially-correlated Rayleigh channel corresponding to user k is a complex normal random vector, with zero mean and spatial correlation matrix $\mathbb{E}\{\mathbf{h}_k \mathbf{h}_k^H\} = \mathbf{R}_k$, $\mathbf{h}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_{N_T}, \mathbf{R}_k)$ [15,19].

The normalized trace $\gamma = \frac{1}{N_T} \text{tr}(\mathbf{R}_k)$ captures the average channel gain from one of the antennas at the base station to the UE and includes the macroscopic large-scale fading. The Rayleigh distribution of \mathbf{h}_k is motivated by the presence of small-scale fading due to multipath propagation. When the channel gains between different antennas are uncorrelated, the channel model can be referred to as uncorrelated Rayleigh fading. γ can be modeled as:

$$\gamma = \gamma_0 - 10\alpha \log_{10} \left(\frac{D}{D_0} \right) + F, \quad (14)$$

where D is the distance between the transmitter and receiver, α is the path loss exponent, which determines how fast the signal attenuates with distances, and γ_0 determines the median channel gain at a reference distance D_0 . In theoretical studies, those two parameters can be computed with one of the many established propagation models. The only non-deterministic terms is $F \sim \mathcal{N}(0, \sigma_{sf}^2)$ and represents the shadow fading that creates log-normal random variations around the mean value, $\gamma_0 - 10\alpha \log_{10} \left(\frac{D}{D_0} \right)$. The shadow fading can be viewed as a model of blockage from large obstacles. The variance σ_{sf}^2 determines how large the variations are.

Realistic performance assessment of MIMO systems requires the use of a channel model that reflects the main characteristics of large antenna arrays such as the array geometry, the correlation between the channel responses of different antennas and the physical location and orientation of BSs and UEs. In [15], a correlation-based model was proposed, where the channel responses were all Gaussian distributed with zero mean and entirely defined through the correlation matrices. Having an intuitive structure, the correlation model allowed the subspaces of the correlation matrices to be parametrized by the azimuth angles of the UEs. This makes it easy to determine if two UEs are spatially separable by comparing their respective angles.

Since all the multipath components are assumed to arrive with the same delay, the model is a frequency flat-fading channel model. This limitation is less significant, as we focus only on communication over a coherence block, where the channel is assumed to be constant. For these reasons, in the context of user clustering for MU-MIMO, the model was employed in several publications like [13,19].

To model the small-scale fading, each path from the antenna to the user is assumed to represent a superposition of N_p multipaths (N_p is a large number). It is assumed that the macro base station is elevated and that the scatterers only surround the user, and each multipath is a plane wave, leaving from the uniform linear spaced array at an angle $\bar{\varphi}_n$. The equivalent channel of each user becomes:

$$\mathbf{h} = \sum_{n=1}^{N_p} \mathbf{a}_n, \tag{15}$$

where the channel vector of the n th path towards the N_T antenna array is:

$$\mathbf{a}_n = g_n \left[1 e^{2\pi j d \sin(\bar{\varphi}_n)} \dots e^{2\pi j d (N_T-1) \sin(\bar{\varphi}_n)} \right]^T, \tag{16}$$

where \mathbf{X}^T denotes the transpose of an array \mathbf{X} , g_n is a complex number that models the gain and phase of multipath n , and d is the array antenna spacing (measured in wavelengths).

It is assumed that the angles $\bar{\varphi}_n$ are random variables with the Probability Density Function (PDF) $f(\bar{\varphi})$ and g_n are zero mean random variables with variance $\mathbb{E} \{ |g_n|^2 \}$. The variance is the n th multipath component average gain, and the total gain of the multipaths is $\gamma = \sum_{n=1}^{N_p} \mathbb{E} \{ |g_n|^2 \}$.

Using the multidimensional central limit theorem, $\mathbf{h} \rightarrow \mathcal{N}_{\mathbb{C}}(\mathbf{0}_{N_T}, \mathbf{R})$ when $N_p \rightarrow \infty$, where the convergence is in its distribution and has a spatial correlation matrix $\mathbf{R} = \mathbb{E} \{ \sum_n \mathbf{a}_n \mathbf{a}_n^H \}$. This is the reason why the channel model is a correlated Rayleigh fading model.

The element $r_{l,m}$ of matrix \mathbf{R} is defined as:

$$\begin{aligned} r_{l,m} &= \sum_{n=1}^{N_p} \mathbb{E} \{ |g_n|^2 \} \mathbb{E} \left\{ e^{2\pi j d (l-1) \sin(\bar{\varphi})} e^{-2\pi j d (m-1) \sin(\bar{\varphi})} \right\} \\ &= \gamma \int e^{2\pi j d (l-m) \sin(\bar{\varphi})} f(\bar{\varphi}) d\bar{\varphi} \end{aligned} \tag{17}$$

As $r_{l,m}$ depends on the $l - m$ difference and not on each individual value, \mathbf{R} is a Toeplitz matrix.

It is assumed that the multipath components are originating from a scattering cluster around the user [due to the lack of scatterers around the Base Station (BS)], located at an angle $\bar{\varphi} = \varphi + \delta$. φ is a deterministic nominal angle, and δ is a random variable with a standard deviation from the nominal angle given by σ_φ .

Depending on how δ is modeled, several correlation models can be found in the literature. The standard deviation $\sigma_\varphi > 0$ is measured in radians and is called the Angular Standard Deviation (ASD). A usual value for urban environments is $\sigma_\varphi > 0$ (10°), but smaller values are expected in rural areas or hills.

When δ follows a Gaussian distribution $\delta \sim \mathcal{N}(0, \sigma_\theta^2)$, an approximate closed-form expression that reduces the complexity of simulations can be used for $r_{l,m}$, if σ_θ is small:

$$r_{l,m}(\theta) = \gamma e^{2\pi j d(l-m) \sin(\theta)} e^{-\frac{\sigma_\theta^2}{2} [2\pi d(l-m) \cos(\theta)]^2}. \tag{18}$$

Accounting for the spatial channel correlation model above, Equation (12) becomes:

$$\mathbb{V} \left\{ \frac{\mathbf{h}_i^H \mathbf{h}_k}{\sqrt{\mathbb{E}\{|\mathbf{h}_i|^2\} \mathbb{E}\{|\mathbf{h}_k|^2\}}} \right\} = \frac{\text{tr}[\mathbf{R}(\theta_i) \mathbf{R}(\theta_k)]}{N_T^2}. \tag{19}$$

Figure 5 plots the metric defined in Section 3.1 for a uniform linear antenna array of size N_T , with the elements placed at half of wavelength and θ_k the angle of the reference user, set at 0° . Over a certain interval, e.g., $\Delta\theta = \theta_i - \theta_k, \Delta\theta \in [-90^\circ, 90^\circ]$, the channel separation metric has a monotone variation with the angle between two co-scheduled users.

Assuming a precoder requires a 0.1 channel separation ($\epsilon = 0.1$), Figure 5 shows that this is achieved for any inter-user angle, when the number of transmit antennas (N_T) is greater than or equal to 32. For 16 antennas, the angle required between two co-scheduled users is 20° and increases up to 70° for the two transmit antennas case.

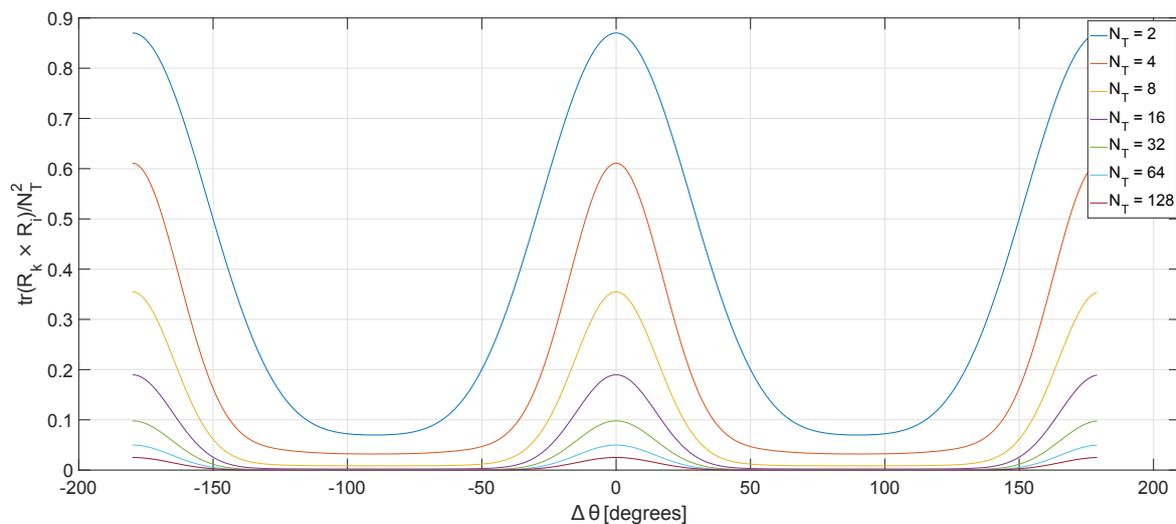


Figure 5. Channel separation as a function of the angle between user k and i , for δ following a Gaussian distribution.

4. User Grouping

One of the objectives of MU-MIMO technology is to send independent data streams to different users. This implies that only a subset of the transmitted data symbols is useful for each co-scheduled user. In the MU-MIMO context, the user grouping is the task of forming a subset of users, according to a spatial compatibility metric. It can be shown that there is a correspondence between the precoding performance and the user grouping technique [9]. The scheduling of spatially-compatible users is crucial to suppress inter-user interference, whereas refining the power allocation and precoding weights requires less computational effort.

Most of the scheduling algorithms in the literature focus on constructing sets of users with orthogonal or semi-orthogonal MIMO channels [20]; however, the optimal solution for the user

grouping involves an extensive search of all possible combinations. The complexity of the approach grows exponentially with the number of users and is not feasible for a moderate number of users.

Scheduling algorithms based on spatial clustering can improve the overall MU-MIMO performance by adjusting the parameter ϵ defined in Section 3.1, according to the number of users and the Signal-to-Noise Ratio (SNR). However, the optimum value is a deployment parameter and depends on the number of transmit antennas, the number of users, the precoder type, and is usually calculated through simulations.

Based on the above observations, in this section we propose a user grouping algorithm, whose aim is to cluster users that have similar directions in Euclidean space. Therefore, their channels exhibit a spatial correlation higher than ϵ and, consequently, cannot be co-scheduled in MU-MIMO through linear precoding schemes, as they may significantly interfere with each other.

To optimize the performance of the spatial clustering, K-means clustering algorithm [21] with a distance metric based on angles between users, is proposed. The algorithm groups the users into N clusters, such that each user belongs to the cluster whose center is the closest (forms the smallest angle from the serving sector perspective), as described in Algorithm 1.

Algorithm 1 K-means clustering.

```

1: for  $N = 1$  to  $K$  do
2:   repeat
3:     Create  $N$  random cluster centers,  $c_n(i)$ 
4:     Compute  $K \times N$  angles for each user to each cluster center
5:     Assign each user  $k$  to the closest cluster  $n$ 
6:      $n = \arg \min_{n \in N} (\angle(k, c_n(i)))$ 
7:     Recompute the centers of the clusters
8:      $c_n(i+1) = \text{Mean}_{k \in n} (\angle(k, c_n(i)))$ 
9:     Stop when the centers do not vary by more than  $err$  percent
10:  until  $\max_{n \in N} (c_n(i) - c_n(i-1)) \leq err$ 
11:  Find the maximum angle between any two users of the same cluster
12:   $angle_{max} = \max_{n \in N} \left( \max_{k_i \in n, k_j \in n} (\angle(k_i, k_j)) \right)$ 
13:  Stop when the users are close enough
14:  if  $angle_{max} \leq \Delta\theta$  then
15:    return  $N$ , cluster assignments
16:  end if
17: end for

```

Although this approach guarantees that users of the same cluster cannot be co-scheduled through linear precoders, it does not provide any information on the channel separation between users of different clusters. To decide whether users from different clusters can be co-scheduled, one can use the angles between the centers of clusters, instead of the angles for the entire set of users. The set of co-scheduled clusters should not include any cluster that has an angle smaller than $\Delta\theta$, and its size should be less than or equal to the number of transmit antennas. The method of finding the optimum sets of co-scheduled clusters is another optimization problem and is not the purpose of this article.

To further reduce the computational burden, once the sets are created, the algorithm will schedule in the same time-frequency resource a subset of users in each cluster of the set through a non-linear precoding scheme. The user selection within each cluster is random; however, this could result in co-scheduling users with bad channel conditions, which would degrade the overall

system performance. Employing more complex selection methods like semi-orthogonal or greedy user selection, as detailed in [4,20,22], would have severely increased the convergence time and the complexity of the algorithm, but was not in the scope of this article. In the context of a hybrid precoding scheme, in order to perform intra-cluster user selection, we would require knowing the effective channel matrix (including the inter-cluster linear precoding). This is conditioned on the users selected in other clusters. As such, the number of combinations is much higher and requires a search over the entire set of users, not only over those of the same cluster.

Note that after spatial clustering, the number of users in each cluster is not the same. Users that have fewer neighbors within their clusters could get higher throughput as they will be co-scheduled more often. For that, this approach is an overlapping user scheduling algorithm [22].

To evaluate the complexity of the spatial clustering algorithm compared to the state-of-the-art solutions, the number of FLOPs is represented for various distance metrics and clustering algorithms, in Figure 6. It can be noted that the complexity of the algorithms depends on the number of users K , the number of antennas N_T , the number of clusters N , and the chosen distance metric. For a fixed number of clusters, $N = 4$ and $N_T = 16$ antennas, the proposed algorithm has better performance than the two-stage precoder with chordal distance (two-stage chordal), which is $\mathcal{O}(K^{N+1} \times N)$ [23], compared to $\mathcal{O}(K^{N+1} \times N \times (N_T^2 + 2N_T^3))$ [5]. In addition, as the proposed metric remains in the Euclidean space, the K-means implementation based on Lloyd’s algorithm is optimized and has an $\mathcal{O}(KNI)$ complexity [24]. Fixing the number of iterations needed until convergence at $I = 50$, the optimized K-means also outperforms the ZF precoding with User Selection (ZF-US), which has an $\mathcal{O}(KN_T^5)$ complexity [22]. The number of iterations will further decrease when the users already have a cluster structure, which is the case of a typical user distribution, as shown in Section 6.

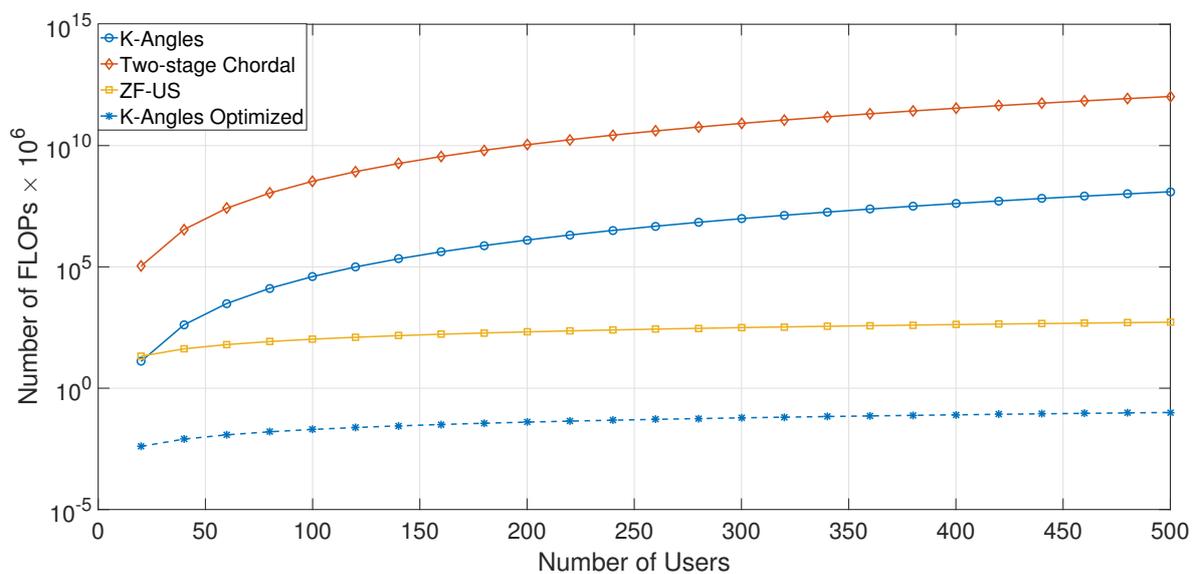


Figure 6. User clustering complexity for various distance metrics and grouping algorithms. US, User Selection.

5. System Model

We consider the downlink of a single-cell MU-MIMO system, where a base station with a uniform linear array of N_T antennas simultaneously transmits data to K single-antenna users. \mathbf{x}_p is the vector of precoded symbols using various precoding schemes p . The received signal vector of the users, \mathbf{y} , becomes:

$$\mathbf{y} = \mathbf{H}\mathbf{x}_p + \mathbf{n}, \tag{20}$$

where $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_K]^T$ is the channel matrix, with \mathbf{h}_k being the channel vector for the k th user. The channel vector of each user is generated from a Rayleigh distribution, and the gains from different antennas are spatially correlated using the model introduced in Section 3.2, such that $\mathbf{h}_k \in \mathbb{C}^{K \times N_T} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_{N_T}, \mathbf{R}_k)$. \mathbf{R}_k is the spatial correlation matrix and $\mathbf{n} = [n_1, \dots, n_K]^T \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma_n^2 \mathbf{I}_K)$ is the vector of AWGN samples of the K users of variance σ_n^2 .

In Figure 7, the system model is illustrated in the context of hybrid precoding. The user grouping algorithm proposed in Section 4 requires the AoA of each user to group the users into clusters. Once the users are clustered, a set of selected users is precoded in each cluster by a THP block. The set of selected clusters is then fed to the BD block to compute the inter-group precoding weights. Both BD and THP precoders require full CSI for the calculation of precoding weights.

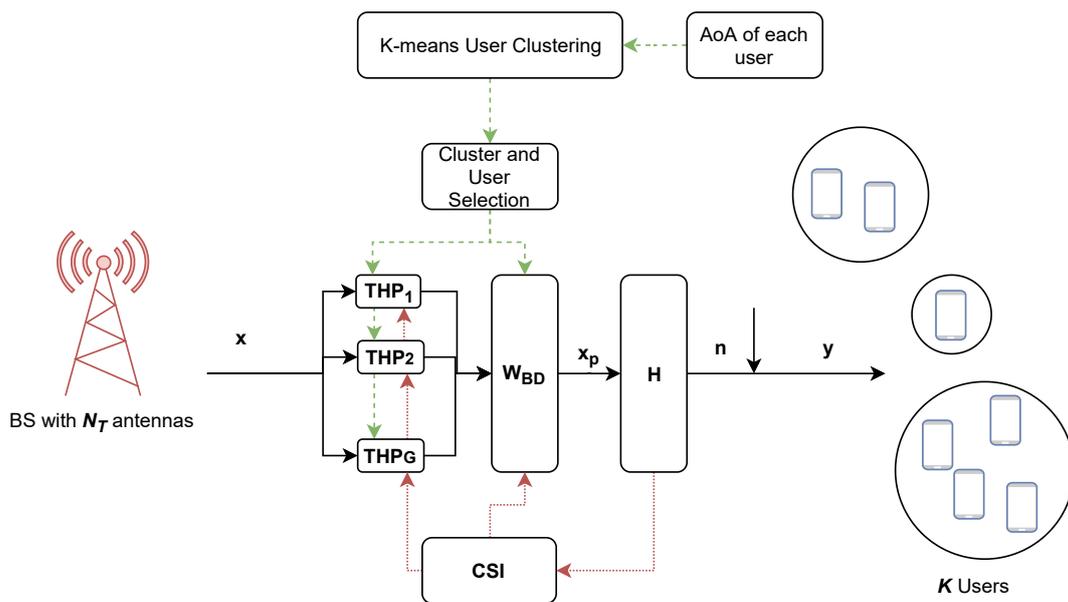


Figure 7. Hybrid precoding with user clustering.

With a correlation-based propagation model assumption, the proposed clustering algorithm does not require full CSI. It only relies on the Angle of Arrival (AoA) to group users together based on their channel correlation. This is already included in modern communication standards like 5G NR, as they are required in beam management procedures, for instance. In our modeling, the angle of each user is assumed to be known at the base station and is computed directly based on its 3D coordinates.

The channel is assumed to be constant during a coherence block and perfectly known at the base station. It is assumed to change independently from one block to another, as a stationary ergodic random process. This assumption does not constrain the applicability of the problem. In the modern wireless communication systems, the Orthogonal Frequency Division Multiplex (OFDM) technology (or similar) is applied, along with a dense grid of pilot reference signals. Temporal variations are also handled through pilot signals and channel estimation. In addition, the channel model is spatially consistent; that is, the channel statistics for a given location are always the same and do not depend on the simulation runs.

The scenarios with users having very high mobility could limit the applicability of the clustering algorithm; however, in the evaluated results, we assumed the users had limited mobility. This is also in line with the distribution used for performance evaluation, where the traffic is focused mostly indoors.

Users with a full-buffer traffic model are distributed within each cell across a 2-km radius. The 3D traffic spatial distribution used was created in the radio planning and optimization software Planet [25], based on geolocated call traces and social media events, in the urban area of Ottawa.

The sum data rate is evaluated with respect to the energy per information bit, E_b , divided by the one-sided noise power spectral density, N_0 , $E_b/N_0 \triangleq P_{TX} / [K\sigma_n^2 \log_2(M)]$, where $P_{TX} = \mathbb{E} \{ \mathbf{s}_x^H \mathbf{s}_x \}$ is the average total transmit power and M is the order of the QAM modulated data symbol.

6. Simulations

This section evaluates the performance in terms of the average sum rate and computational complexity of the presented precoding schemes. Spatio-temporally-correlated channels and perfect channel estimation were assumed. Monte Carlo simulations were performed in a single-cell MU-MIMO system with $N_T = 16$ and $K = 16$ single-antenna users. A 16 QAM modulation was used throughout the simulations. We assumed the channel changed non-coherently from one block to another.

Figure 8 evaluates the sum rate performance of the RZF and THP precoding schemes as a function of E_b/N_0 and the inter-user angular separation, when the users were equally spaced with an angle $\theta = \theta_{\min}$. The performance of both schemes decreased as the users became closer to each other; however, THP always outperformed RZF since it employs a more sophisticated successive interference cancellation technique. For example, to reach the 50 b/s/Hz, RZF requires 17 dB more E_b/N_0 when $\theta_{\min} = 1^\circ$ than when $\theta_{\min} = 5^\circ$. In the same scenario, THP requires just 8 dB more E_b/N_0 . Although these results reflect the performance under equal angle separation, this is the lower bound in a worst case scenario, and similar trends are expected when certain users are separated at different angles.

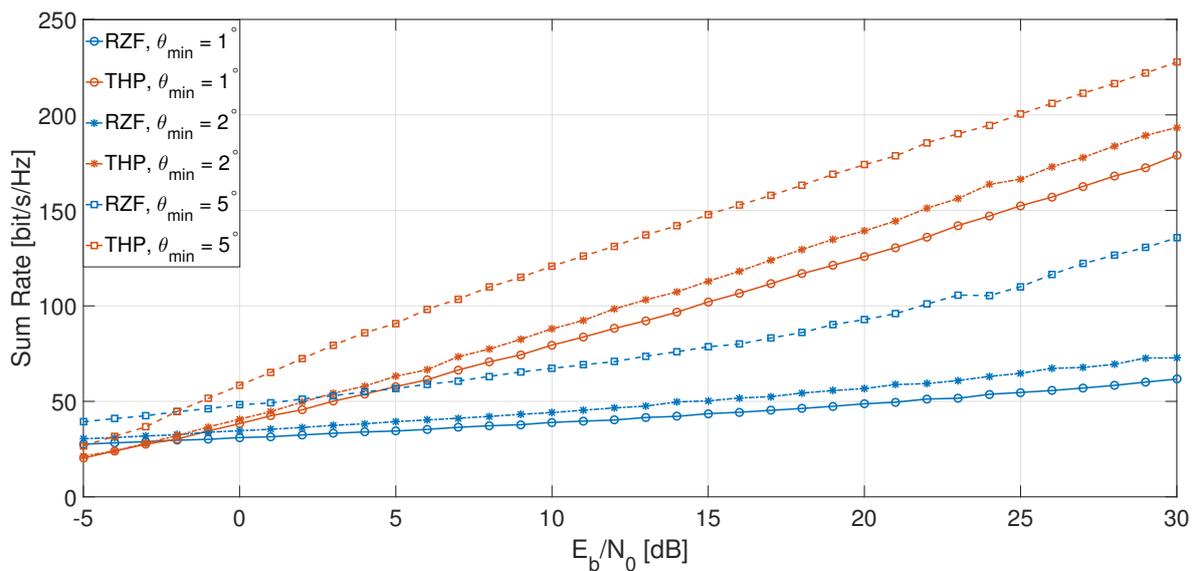


Figure 8. Sum rate vs. E_b/N_0 and θ_{\min} .

To plot the sum rate performance of a hybrid BD-THP precoding scheme in Figure 9, it was assumed that the users could be grouped into four groups located at $\theta_g = [-45^\circ, -15^\circ, 15^\circ, 45^\circ]$. The users within each group were uniformly spread around θ_g and were equally spaced with an angle of 1° . As the minimum angle between any two groups was 30° and between any two users of the same group was 1° , we could apply a linear precoding scheme between groups and non-precoding within the groups. The performance of hybrid BD-THP precoding was better than the RZF precoding, except for low values of E_b/N_0 , as BD did not account for the noise enhancement.

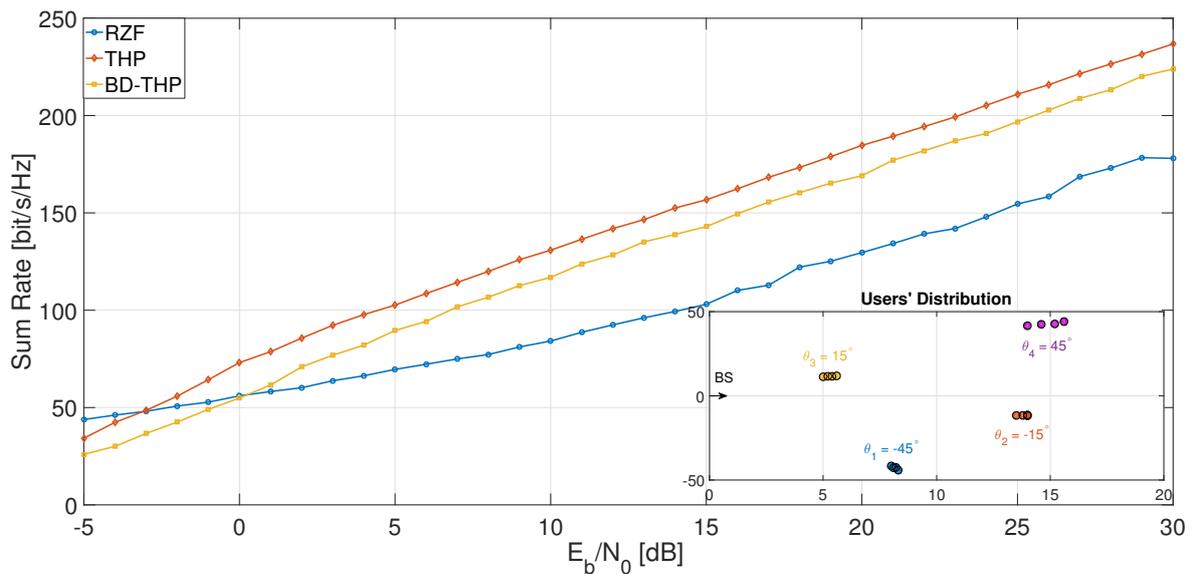


Figure 9. Sum rate vs. E_b/N_0 for $\theta_{\min,g} = 1^\circ$ and $\theta_g = [-45^\circ, -15^\circ, 15^\circ, 45^\circ]$.

In the following, the user grouping algorithm and the performance of precoders are evaluated, accounting for a realistic user spatial distribution.

In Figures 10 and 11, Sector B2 serves users spread using two different traffic distributions, in a 120° azimuth. The first distribution captured the working day traffic from the morning and the evening hours (busy hours) and the second, in between (working hours). Due the mobility of users, the busy hour traffic was spread more than during the working hours, when it was concentrated in hotspots (around office buildings). To limit the performance degradation of linear precoding schemes, the required angular separation between clusters was set to 30° . This resulted in users being grouped into four clusters for the working hour user distribution and in three clusters during busy hours.



Figure 10. Working hour user distribution and clustering for Sector B2 and $\Delta\theta = 30^\circ$.



Figure 11. Busy hour user distribution and clustering for Sector B2 and $\Delta\theta = 30^\circ$.

In practice, the AoA is not always perfectly estimated. For that, in Figures 12 and 13, we evaluated the impact of the AoA estimation error on the clustering algorithm’s performance. The AoA estimation error was modeled as a normal random variable, with zero mean and a variable standard deviation. Because the angles between users were computed by the K-means clustering algorithm, the error was added to the actual coordinates of the users.

In Figure 12, the probability of grouping users into the same clusters (as in the ideal case of perfect AoA estimation) was investigated for different values of the standard deviation error and number of clusters. As expected, as the AoA error increased, the probability of assigning users in the same clusters decreased. The impact was even higher when the number of clusters increased.

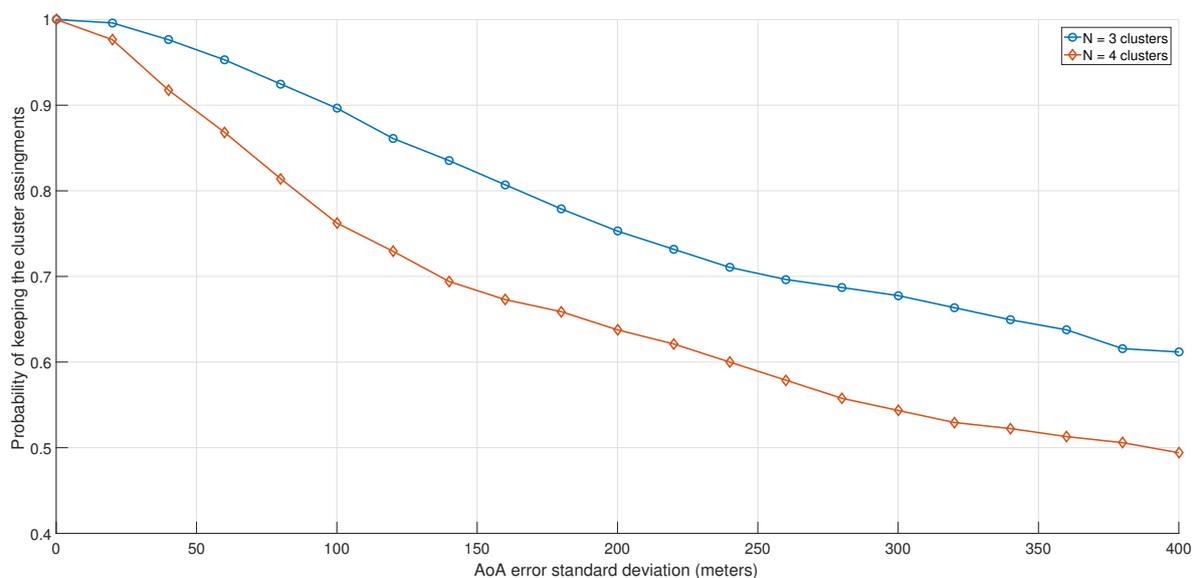


Figure 12. Probability of users keeping the same clusters, as a function of the AoA error.

Figure 13 plots the average number of iterations until convergence as a function of the AoA standard deviation of error and the number of clusters. It can be noted that the convergence was less dependent on the error and slightly increased for higher AoA error values.

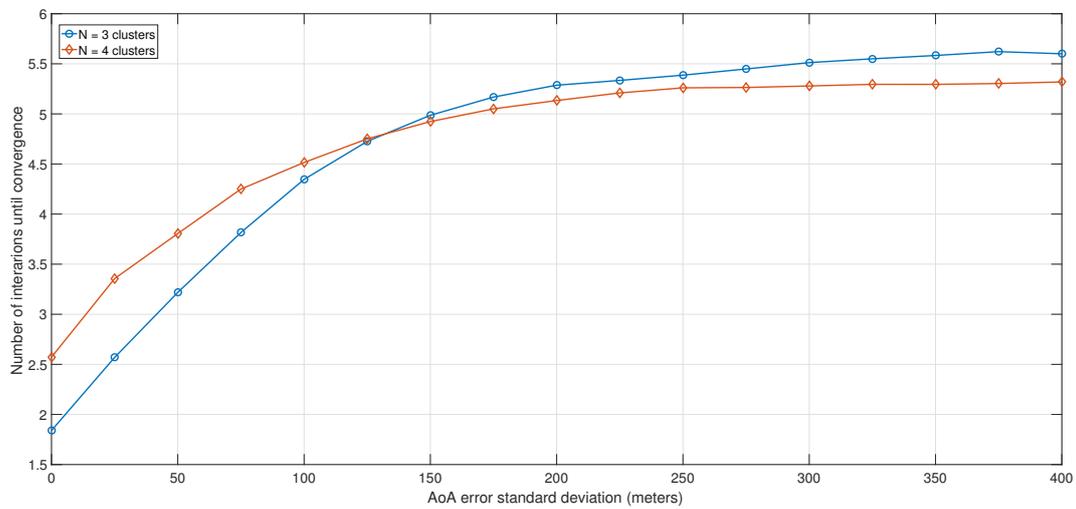


Figure 13. Number of iterations until convergence, as a function of the AoA error.

To plot the sum rate performance of a hybrid BD-THP precoding scheme, in Figure 14, we used the busy hour user distribution above and applied the BD linear precoding between the three clusters. Inside each cluster, 6, 6, and 4 users were randomly selected and precoded through a non-linear THP precoding scheme, as a total of 16 users could be co-scheduled through MU-MIMO. Compared to the state-of-the-art precoding schemes like THP and RZF, the hybrid precoding followed the trend of the THP precoding, although the performance was slightly reduced due to the inferior performance of the linear precoding. It can be shown that even in Rayleigh uncorrelated channels, the linear precoding performance is below the non-linear precoding performance.

In BD-RZF, K-means clustering with a chordal distance metric was used to group the users into a two-stage precoding scheme. BD-RZF scheme had lower performance than BD-THP because, for $K = N_T$, there were not enough degrees of freedom to separate the users in the spatial domain when applying linear precoding techniques. The performance was even worse than of the RZF precoder, due to reducing the effective channel matrix and using statistical CSI to reduce the complexity. This observation is also in line with the findings in [13].

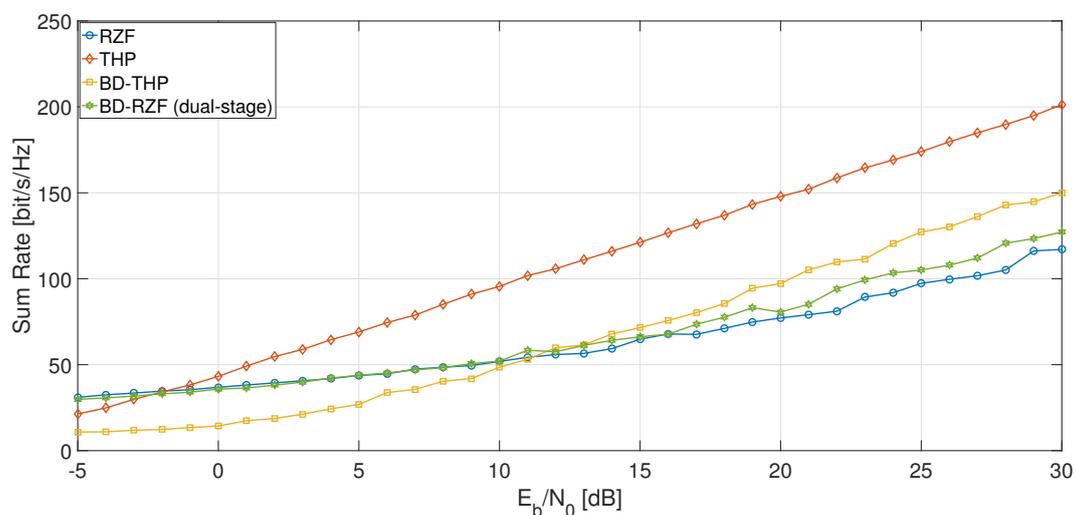


Figure 14. Sum rate vs. E_b/N_0 for busy hour user distribution and random user selection.

7. Conclusions

In this paper, a user grouping algorithm based on a modified K-means spatial clustering was proposed. The technique tried to address the complexity of scheduling spatially-compatible users, as required by different precoding schemes.

It was shown that the user's channels inter-correlation depended on the transmit correlation matrix, for a base station with a uniform linear array of antennas and single antenna users, and that the user's channels inter-correlation depended ultimately on the angles between the users.

The proposed grouping algorithm identified clusters of users where the average angle between users was minimized. Given that in real life, the locations of users follow a distribution with a clustered structure, the proposed approach could be a very efficient way of identifying users with separable channels. In this case, we showed that it had a linear complexity with the number of users.

The performance of various precoding schemes like linear RZF, BD, non-linear THP, and the hybrid, BD-THP, was evaluated, and it was shown that linear precoding schemes, like RZF, were more affected by users with correlated channels than the non-linear ones. For that reason, hybrid precoding is a good tradeoff between complexity and performance.

Finally, the performance of the hybrid BD-THP precoding scheme was evaluated on a realistic user distribution using the proposed clustering algorithm.

Although we showed that the AoA errors had a limited impact on the clustering results, more work is required to evaluate the overall precoding algorithm performance correctly under non-ideal channel estimation and AoA errors.

Author Contributions: Conceptualization, R.-F.T.; methodology, A.-A.E.; formal analysis, C.P.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Chien, T.V.; Bjornson, E. *Massive MIMO Communications, 5G Mobile Communications*; Springer: Basel, Switzerland, 2017; pp. 77–116.
- Castaneda, E.; Silva, A.; Gameiro, A.; Kountouris, M. An Overview on Resource Allocation Techniques for MU-MIMO Systems. *IEEE Commun. Surv. Tutor.* **2016**, *19*, 239–284. [[CrossRef](#)]
- Trifan, R.F.; Paleologu, C. Non-Linear Precoding Performance in Spatio-Temporally Correlated MU-MIMO Channels. In Proceedings of the 2018 International Conference on Communications (COMM), Bucharest, Romania, 14–16 June 2018.
- Adhikary, A.; Caire, G. Joint Spatial Division and Multiplexing: Opportunistic Beamforming and User Grouping. *IEEE J. Sel. Top. Signal Process.* **2014**, *8*, 879–890.
- Xu, Y.; Yue, G.; Mao, S. User Grouping for Massive MIMO in FDD Systems: New Design Methods and Analysis. *IEEE Access* **2014**, *2*, 947–959. [[CrossRef](#)]
- Majumdar, I. Implementation of Block Diagonalization Type Precoding Algorithms for IEEE 802.11ac Systems. In Proceedings of the 2015 Fifth International Conference on Advances in Computing and Communications (ICACC), Kochi, India, 2–4 September 2015.
- Cho, Y.S.; Kim, J.; Yang, W.Y.; Kang, C.G. *MIMO-OFDM Wireless Communications with Matlab*; Wiley: Hoboken, NJ, USA, 2010.
- Spencer, Q.H.; Swindlehurst, A.L.; Haardt, M. Zero-forcing methods for downlink spatial multiplexing in MU-MIMO channels. *IEEE Trans. Signal Process.* **2004**, *52*, 461–471. [[CrossRef](#)]
- Trifan, R.F.; Enescu, A.A. MU-MIMO Precoding Performance Conditioned by Inter-user Angular Separation. In Proceedings of the 2018 International Symposium on Electronics and Telecommunications (ISETC), Timisoara, Romania, 8–9 November 2018.
- Fischer, R.F.H.; Windpassinger, C.A. Improved MIMO Precoding for Decentralized Receivers Resembling Concepts from Lattice Reduction. In Proceedings of the GLOBECOM '03, IEEE Global Telecommunications Conference (IEEE Cat. No.03CH37489), San Francisco, CA, USA, 1–5 December 2003.

11. Hasegawa, F. MU-MIMO performance evaluation of nonlinear precoding schemes. In Proceedings of the 3GPP TSG RAN WG1 Meeting 88b, Spokane, WA, USA, 3–7 April 2017.
12. Classon, B. Non-linear Precoding for Downlink MU-MIMO. In Proceedings of the 3GPP TSG RAN WG1 Meeting 88, Athens, Greece, 13–17 February 2017.
13. Zarei, S.; Gerstacker, W.; Schober, R. Low Complexity Hybrid Linear/Tomlinson–Harashima Precoding for Downlink Large-Scale MU-MIMO Systems. In Proceedings of the 2016 IEEE Globecom Workshops (GC Wkshps), Washington, DC, USA, 4–8 December 2016.
14. Costa, M. Writing on Dirty Paper. *IEEE Trans. Inf. Theory* **1983**, *29*, 439–441. [[CrossRef](#)]
15. Bjornson, E.; Hoydis, J.; Sanguinetti, L. Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency. *Found. Trends Signal Process.* **2017**, *11*, 154–655. [[CrossRef](#)]
16. Herdin, M.; Czink, N.; Ozcelik, H.; Bonek, E. Correlation Matrix Distance, a Meaningful Measure for Evaluation of Non-Stationary MIMO Channels. In Proceedings of the 2005 IEEE 61st Vehicular Technology Conference, Stockholm, Sweden, 30 May–1 June 2005.
17. Wang, B.H.; Hui, H.T.; Yu, Y.T. Maximum volume criterion for user selection of MU-MIMO downlink with multiantenna users and block diagonalization beamforming. In Proceedings of the IEEE Antennas and Propagation Society International Symposium, Toronto, ON, Canada, 11–17 July 2010; pp. 1–4.
18. Godana, B.; Ekman, T. Parametrization Based Limited Feedback Design for Correlated MIMO Channels Using New Statistical Models. *IEEE Trans. Wirel. Commun.* **2013**, *12*, 5172–5184. [[CrossRef](#)]
19. Adhikary, A.; Nam, J.; Ahn, J.Y.; Caire, G. Joint Spatial Division and Multiplexing The Large-Scale Array Regime. *IEEE Trans. Inf. Theory* **2013**, *59*, 6441–6463.
20. Yoo, T.; Goldsmith, A. On the Optimality of Multiantenna Broadcast Scheduling Using Zero-Forcing Beamforming. *IEEE J. Sel. Areas Commun.* **2006**, *24*, 528–541.
21. Lloyd, S.P. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28*, 129–137. [[CrossRef](#)]
22. Tian, R.; Liang, Y.; Tan, X.; Li, T. Overlapping User Grouping in IoT Oriented Massive MIMO Systems. *IEEE Access* **2017**, *5*, 14177–14186. [[CrossRef](#)]
23. Inaba, M.; Katoh, N.; Imai, H. Applications of weighted Voronoi diagrams and randomization to variance-based k-clustering. In Proceedings of the 10th ACM Symposium on Computational Geometry, Stony Brook, NY, USA, 6–8 June 1994.
24. Hamerly, G. Making k-means Even Faster. In Proceedings of the SIAM International Conference on Data Mining, Columbus, OH, USA, 29 April–1 May 2010.
25. Planet. Available online: <https://www.infovista.com/products/planet-network-planning-solutions> (accessed on 18 May 2019).



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).