

Article

# Spatio-Radio Resource Management and Hybrid Beamforming for Limited Feedback Massive MIMO Systems

Hedi Khammari <sup>1</sup>, Irfan Ahmed <sup>2,\*</sup> , Ghulam Bhatti <sup>3</sup> and Masoud Alajmi <sup>1</sup>

<sup>1</sup> Department of Computer Engineering, College of Computers and Information Technology, Taif University, Taif 21974, Saudi Arabia; hkhammari@tu.edu.sa (H.K.); ms.alajmi@tu.edu.sa (M.A.)

<sup>2</sup> Department of Electrical Engineering, Higher Colleges of Technology, Ruwais 12389, UAE

<sup>3</sup> Department of Computer Science, College of Computers and Information Technology, Taif University, Taif 21974, Saudi Arabia; gbhatti@tu.edu.sa

\* Correspondence: iahmed3@hct.ac.ae

Received: 5 August 2019; Accepted: 17 September 2019; Published: 20 September 2019



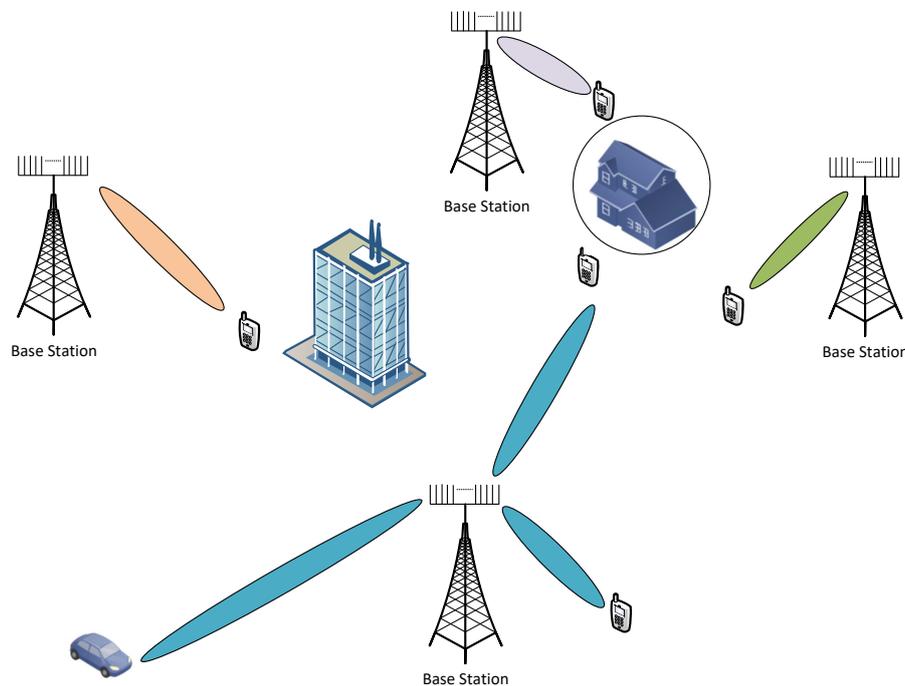
**Abstract:** In this paper, a joint spatio–radio frequency resource allocation and hybrid beamforming scheme for the massive multiple-input multiple-output (MIMO) systems is proposed. We consider limited feedback two-stage hybrid beamforming for decomposing the precoding matrix at the base-station. To reduce the channel state information (CSI) feedback of massive MIMO, we utilize the channel covariance-based RF precoding and beam selection. This beam selection process minimizes the inter-group interference. The regularized block diagonalization can mitigate the inter-group interference, but requires substantial overhead feedback. We use channel covariance-based eigenmodes and discrete Fourier transforms (DFT) to reduce the feedback overhead and design a simplified analog precoder. The columns of the analog beamforming matrix are selected based on the users' grouping performed by the K-mean unsupervised machine learning algorithm. The digital precoder is designed with joint optimization of intra-group user utility function. It has been shown that more than 50% feedback overhead is reduced by the eigenmodes-based analog precoder design. The joint beams, users scheduling and limited feedbackK-based hybrid precoding increases the sum-rate by 27.6% compared to the sum-rate of one-group case, and reduce the feedback overhead by 62.5% compared to the full CSI feedback.

**Keywords:** hybrid beamforming; massive MIMO; resource allocation

## 1. Introduction

The scarcity of available frequency band for wireless communications has led to the inclusion of millimeter Wave (mmWave) frequencies in cellular communications. This has opened the doors for massive multiple-input multiple-output (MIMO) systems. Due to high transmission frequencies, fabrication of large number of antennas with small form factor has become possible. MmWave band has inherent hindrances, like, high path-loss and absorption-loss. It has been known that MIMO systems advantages (spatial multiplexing or diversity gain) are scaled-up with the number of antennas. In summary, one can enjoy the benefits of the large bandwidth available at mmWave frequencies by combating high path and absorption losses with massive MIMO directional beamforming. Future mmWave massive MIMO-based cellular networks will be as shown in Figure 1. Due to the high pathloss on one hand and high directional gain on the other hand, the inter-cell interference and cell boundaries will become meaningless. The fixed area size cell boundaries of traditional cellular will probably no longer exist in the future mmWave massive MIMO systems. Narrow beams can serve

distant user equipment (UE) without interfering other UEs provided that there is no obstacle between BS and intended UE, whereas a closely located UE may be deprived of connection due to the obstacles.



**Figure 1.** MmWave massive MIMO based cellular system.

The cost of massive MIMO is in terms of excessive feedback overhead for channel estimation along with the hardware complexity of RF chains (increased number of radio-frequency (RF) chains). The feedback overhead has been tackled separately for frequency division duplex (FDD) and time division duplex (TDD) systems. In FDD systems, the uplink channel estimation consists of fewer overheads compared to the downlink channel estimation, because generally, the number of transmit antennas  $N_t$  is larger than the number of users  $K$ , and the number of receive antennas per user  $n_k$  ( $N_t \gg K$  and  $N_t \gg n_k$ ). The most common technique to reduce the downlink channel estimation overhead is joint spatial division multiplexing (JSDM) [1]. The JSDM uses two-stage precoding: second order channel statistics (covariance)-based user grouping and the traditional MU-MIMO linear precoding (zero-forcing) for the inter-user interference mitigation based on the low-dimensional effective channel. In TDD, only uplink channel estimation is done and the downlink channel estimates are obtained by the transpose of the uplink channel using the channel reciprocity principle. The TDD massive MIMO systems suffer from pilot contamination when the BS receives non-orthogonal pilot signals from the neighboring cells. This pilot contamination degrades the channel estimation and hence, affects both uplink combining and downlink precoding.

In traditional MIMO systems, a separate RF chain (analog-to-digital converter/digital-to-analog converter, serial-to-parallel/parallel-to-serial converter, up/down converter etc) is required for each antenna, but the high power consumption makes it infeasible for the case of massive MIMO systems. Hybrid beamforming technique resolves this problem by dividing the precoding/combining into baseband digital processing and RF analog processing. The hybrid precoding and combining offer extra degrees of freedom in space domain with a large number of antennas and analog beamforming [2]. The hybrid beamforming can be realized by using MU-MIMO precoding as baseband digital precoding and the statistical channel state information-based pre-beamforming as RF analog precoding. This limited feedback (due to average CSI) configuration is particularly suited for massive MIMO mmWave systems with a large number of antennas but relatively small number of RF

chains [3]. It has been shown [4], that the covariance-based limited feedback works well for mmWave massive MIMO systems, where the number of users is small with respect to the number of BS antennas and the channels are formed by a few multi path components (MPCs) with small angular spread.

Limited work has been done on the joint multiuser massive MIMO resource allocation and hybrid beamforming design. Although mmWave massive MIMO system has a potential of tremendous increase of spectral efficiency. However, the cost and power consumption of power-hungry radio frequency chains (analog-to-digital converter (ADC)/digital-to-analog converter (DAC), parallel to serial converter, serial to parallel converter, up converter/down converter) make it impractical to build a complete RF chain for each antenna. A promising solution to this problem is hybrid beamforming, where the precoder at the transmitter is divided into two parts: analog precoder and digital precoder. The analog precoder (usually a network of phase shifters) at the RF stage reduces the number of RF chains required for the digital precoder. In order to configure these precoders, the transmitter requires channel state information in the form of uplink feedback from users, but in the presence of massive antennas, this feedback becomes a huge load on the wireless uplink, especially in FDD mode. JSDM [4] is a technique used to reduce the feedback overhead. It uses slowly varying average channel statistics to implement the analog precoder; then, the digital precoder is realized by using a low-dimensional effective channel. Till now, different variants of the JSDM have been proposed. Li et al. [5] generalize the JSDM scheme to support non-orthogonal virtual sectorization and with multiple RF chains at both link ends. It uses the Kronecker channel model to decouple the transmit and receive beamforming. Under this channel, the analog beamformer is obtained by stacking strongest eigenbeams of the channel covariance matrix and then the digital beamformer is based on a weighted minimum mean squared error (MMSE) with effective channel. However, the Kronecker model does not characterize the mmWave channel where transmitter and receiver have coupling effects due to highly directional transmission. In [6], the authors apply JSDM using a geometrical channel model and find hybrid precoder and combiner at transmitter and receivers, respectively. Hybrid beamforming with switches (HBwS) has been introduced in [7], where,  $L \times N_t$  analog beamformer is controlled by  $N_{RF} \times L$  instantaneous CSI based switches.  $N_t$  is the number of transmit antennas,  $N_{RF}$  is the number of RF chains, and  $N_{RF} < L < N_t$ . Another switch-based analog beamforming is proposed in [8] but it requires instantaneous CSI for both switching network and the phase shifter network. Also it contains  $L = N_t$ . The JSDM implementation also requires the training in the downlink to estimate the channel covariance matrix. Most of the work assumes that the CSI is known at both ends. In [9], authors consider the joint optimization of the training resource allocation and channel-statistics-based analog beamformer design by using user centric virtual sectorization. There are different structures for the phase shifter-based analog beamformer, namely, fully connected, sub-connected, and dynamically connected [10]. Park [11] investigate JSDM with these analog beamformer architectures. The dynamic architecture gives better result at the cost of added complexity. In [12], authors propose a hybrid beamforming method with unified analog beamformer by Subspace Construction (SC) based on partial CSI in massive MIMO OFDM system. In [13], statistical CSI based analog beamformer uses regularized block diagonalization to mitigate the inter-group interference and instantaneous CSI based digital beamformer utilizes the weighted MMSE to suppress intra-group interference. Jiang et al. [14] jointly optimize the user selection and beam selection during analog beamforming design. They use Lyapunov-drift optimization framework to obtain the optimal solution. Their work only focuses on the design of statistical CSI based analog precoder and user/beam selection. Our previous work [15] on resource allocation for transmit beamforming develops digital and analog precoders which maximize the sum rate with total power and desired number of RF chains constraints. The provided solutions require full instantaneous CSI at the transmitter and receiver, which, in case of the massive MIMO, consists of large number of pilot transmission in downlink and channel information feedback in the uplink. In this work we exploit the channel similarities by grouping (K-Mean machine learning) the users based on the location information. Low complexity DFT matrix based analog precoder is

derived using statistical CSI. This greatly reduces the feedback overhead for the design of zero-forcing digital precoder.

Machine learning (ML) applications for the physical layer of wireless communication systems have been widely reported in [16]. Most of the conventional transmitter and receiver blocks can be replaced by an ML-based auto encoder as suggested by the authors. The large number of antennas in massive MIMO leads to the challenging issue of channel estimation in mmWave communications. A common practice in TDD massive MIMO systems is to utilize the channel reciprocity to get the downlink CSI from uplink channel information estimates. However, in FDD, the channel reciprocity is not applicable and the downlink CSI estimation is very difficult. The downlink channel estimation is known to be hampered by the pilot contamination effect (user to base-station). The quality of channel estimates is deteriorated by the mutual interference caused by the non-orthogonal pilots in a cell. In [17], a supervised learning-based pilot decontamination scheme for massive MIMO uplink is reported. In the proposed ML-based solution, the users' locations in all cells and the pilot assignments stand for the input features and output labels, respectively. In [18], a deep learning network CsiNet is used to learn the CSI-to-codeword transformation (codebook approach is usually adopted to reduce the feedback overhead) at users' terminals and inverse CsiNet at base-station. The authors of [19], suggest a learning-based antenna selection for massive MIMO systems. It uses a multiclass K-NN and support vector machine (SVM) for data-driven optimal antenna selection. Wang et al. [20] employs K-nearest neighbor ( $K - NN$ ) supervised learning for the  $N$  beams allocation among  $K$  users. In [21], a reinforcement learning based framework for radio resource management in radio access networks has been proposed. In our previous work [22], we used neural networks to reduce the execution time of the computationally intensive resource allocation part of the joint resource allocation and hybrid beamforming design in [15]. However, in this work, we use K-mean based unsupervised machine learning scheme to group the users based on their spatial locations. To the best of our knowledge, there is no research work that jointly consider the spatio-radio resource management and the hybrid beamforming in massive MIMO systems.

In this work, we use spatial channel covariance matrices for the analog beamforming design. We also consider the users to RF beam mapping. This mapping requires channel state information and a search over all possible beam combinations at the base-station. This search is exponential in the number of users [23]. Due to this exponential increase in complexity, we use DFT-based eigenmode beams with RF switches.

**Contribution:** In this paper, we develop joint spatio-radio resource and hybrid precoding algorithms for limited feedback wideband massive MIMO systems. The contributions of this paper are summarized as follow.

- First, we consider the problem of joint hybrid precoder design with limited feedback and user-beam selection to maximize the sum proportional rate under the total power constraint. The formulated mixed integer programming problem is then transformed to the relaxed-convex optimization problem.
- Second, a low complexity suboptimal solution is provided for the optimization problem. The algorithm generates the analog beamforming matrix, digital beamforming matrix, and the set of users in each group. The DFT/eigenmodes-based analog beamforming is formed using limited statistical CSI feedback from the users. Then, the digital precoder design with users selection is done iteratively.
- Finally, we develop a K-Mean algorithm based unsupervised machine learning scheme for users grouping. These users groups are used to form the limited feedback (statistical channel state information) based analog beamforming matrix. The proposed machine learning based analog beamforming along with the zero-forcing digital precoding and user scheduling gives better performance than the DFT/eigenmodes-based solution.

The rest of the paper is organized as follows. System, signal, and channel model along with the problem formulation are described in Section 2. Section 3 introduce the relaxed-convex transformation of the formulated mixed integer optimization problem. Suboptimal solution to the joint resource allocation and hybrid beamforming based on eigenmodes and discrete Fourier transform is given in Section 4. Section 5 proposes machine learning based users grouping and beam selection for joint optimization problem. Simulation results are given in Section 6, followed by the conclusions in Section 7.

*Notations:* Bold upper and lower case letters denote vectors and matrices, respectively. The notations  $\mathbf{X}^{-1}$ ,  $\mathbf{X}^\dagger$ ,  $\mathbf{X}^T$ ,  $\mathbf{X}^H$ , and  $tr(\mathbf{X})$  denote the inverse, pseudo-inverse, transpose, Hermitian transpose, and trace of a matrix  $\mathbf{X}$ .  $vec\{\cdot\}$  is a vector operator,  $diag\{x_1, \dots, x_n\}$  is diagonal matrix, and  $\otimes$  is the Kronecker product.  $\|\cdot\|_F$  denotes the Frobenius norm. The  $n \times n$  identity matrix is denoted by  $\mathbf{I}_n$ .  $\mathbb{E}\{\cdot\}$  represents the expectation with respect to the random variable within the brackets.

## 2. System Model

Consider a FDD MU-MIMO downlink system where a base station (BS) with  $N_t$  antennas is located at the cell center and transmits to  $K$  single antenna users as shown in Figure 2. There are  $G$  groups of users such that the group  $g \in \mathcal{G} = \{1, \dots, G\}$ . Each group contains  $K_g$  users.

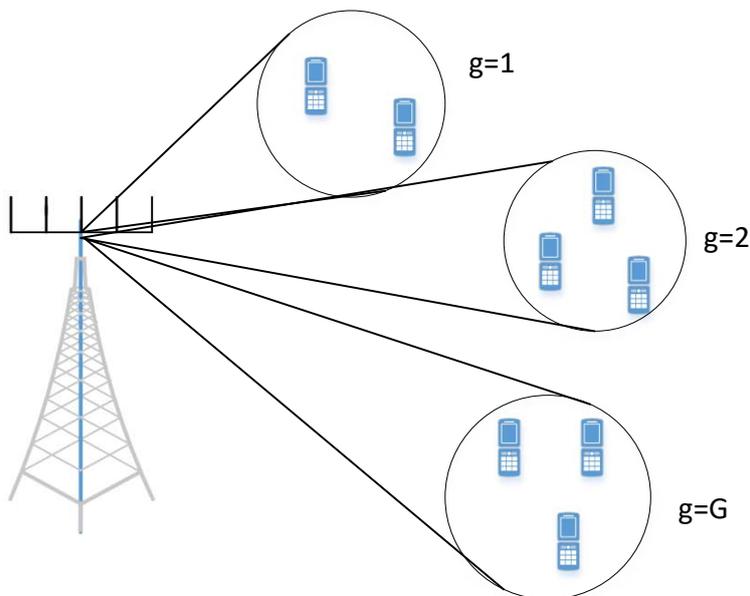


Figure 2. System Model.

Assume that the BS and users have the knowledge of the channel. We consider multi-carrier OFDM transmission with narrow-band block-fading channel. The BS is equipped with  $N_t$  antennas in linear antenna array (ULA) configuration. The information signal block  $\mathbf{S} \in \mathbb{C}^{K \times N_f}$  at the input of the BS transceiver for the user  $k$  is given as

$$\mathbf{s}_k = [s_1, s_2, \dots, s_{N_f}], \quad \forall k, \tag{1}$$

and for the subchannel  $n$ ,

$$\mathbf{s}_n = [s_1, s_2, \dots, s_K]^T, \quad \forall n, \tag{2}$$

where  $N_f$  and  $N_s$  are the number of subchannels and the number of symbols per subchannel, respectively. In a subchannel  $n$ , the information symbol vector is  $\mathbf{s} \in \mathbb{C}^{N_s \times 1}$ . We assume  $N_s = K$ , such that the transmit signal per subchannel  $n$  satisfying  $\mathbb{E}\{\mathbf{s}_n \mathbf{s}_n^H\} = \frac{P_n}{K} \mathbf{I}_K$ , where  $P_n = P_T / N_f$  is the transmit power per subchannel and  $P_T$  is the total transmit power of the BS. The transmit signal vector

$\mathbf{X}$  is obtained from  $\mathbf{F}^B \mathbf{S}$ , where  $\mathbf{F}^B \in \mathbb{C}^{N_t \times N_s}$  is the precoding matrix. The hybrid beamforming divides the precoding matrix into baseband digital precoding matrix  $\mathbf{F}^{DB} \in \mathbb{C}^{N_{RF} \times N_s}$  and RF analog precoding matrix  $\mathbf{F}^{AB} \in \mathbb{C}^{N_t \times N_{RF}}$ , where  $N_{RF}$  is the number of RF chains as shown in Figure 3. The transmit signal  $\mathbf{X} \in \mathbb{C}^{N_t \times N_f}$  is given by

$$\begin{aligned} \mathbf{X} &= \mathbf{F}^B \mathbf{S} \\ &= \mathbf{F}^{AB} \mathbf{F}^{DB} \mathbf{S} \end{aligned} \tag{3}$$

Also, the precoding matrix must satisfy

$$\begin{aligned} \mathbb{E}\{tr(\mathbf{X}\mathbf{X}^H)\} &\leq P_T \\ \mathbb{E}\{tr(\mathbf{F}^{AB} \mathbf{F}^{DB} \mathbf{S} \mathbf{S}^H \mathbf{F}^{DBH} \mathbf{F}^{ABH})\} &\leq P_T \\ tr(\mathbf{F}^{AB} \mathbf{F}^{DB} \mathbb{E}\{\mathbf{S}\mathbf{S}^H\} \mathbf{F}^{DBH} \mathbf{F}^{ABH}) &\leq P_T \end{aligned} \tag{4}$$

since  $\mathbb{E}\{\mathbf{S}\mathbf{S}^H\} = \frac{P_T}{N_f K} \mathbf{I}_{KN_f}$ , therefore,

$$tr(\mathbf{F}^{AB} \mathbf{F}^{DB} \mathbf{F}^{DBH} \mathbf{F}^{ABH}) \leq N_f K \tag{5}$$

The transmit signal in subchannel  $n$  is  $\mathbf{x}_n \in \mathbb{C}^{N_t \times 1}$ . Thus, the received signal vector  $\mathbf{y}_n \in \mathbb{C}^{K \times 1}$  at  $K$  users in subchannel  $n$  is given by

$$\begin{aligned} \mathbf{y}_n &= \mathbf{H}_n^H \mathbf{x}_n + \mathbf{w}_n \\ &= \mathbf{H}_n^H \mathbf{F}^{AB} \mathbf{F}_n^{DB} \mathbf{s}_n + \mathbf{w}_n \end{aligned} \tag{6}$$

where  $\mathbf{H}_n \triangleq [\mathbf{h}_{1,n}, \dots, \mathbf{h}_{K,n}] \in \mathbb{C}^{N_t \times K}$  is the channel matrix with  $\mathbf{h}_{k,n} = [h_{1,k}, \dots, h_{N_t,k}]^H$  being the channel vector from BS to user  $k$  in subchannel  $n$ ,  $\mathbf{x}_n = \mathbf{F}_n^{DB} \mathbf{F}^{AB} \mathbf{s}_n$ , and  $\mathbf{w}_n \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_K)$  be the additive white Gaussian noise (AWGN) in subchannel  $n$  at the users. The RF beamforming  $\mathbf{F}^{RF}$  is performed in time domain and the same beamforming is applied on all subchannels, whereas, the digital beamforming  $\mathbf{F}_n^{DB}$  is performed in frequency domain on the per subchannel basis [11]. In the  $n^{th}$  subchannel, the  $j^{th}$  UE receives the sum of all transmitted signals for  $K$  UEs over its MIMO channel  $\mathbf{H}_{j,i}$  as

$$y_{j,n} = \sum_{k=1}^K \mathbf{h}_{j,n}^H \mathbf{x}_{k,n} + w_{j,n} \tag{7}$$

where  $\mathbf{h}_{j,n}$  is the  $N_t \times 1$  channel vector. We denote the rank of the channel matrix  $\mathbf{H}_{j,n}$  by  $r_{j,n}$ , where  $0 \leq r_{j,n} \leq \min(K, N_t)$ ,  $\forall n$ . In matrix form, the above equation is given as

$$\mathbf{y}_{j,n} = \mathbf{h}_{j,n}^H \mathbf{x}_n + \mathbf{w}_{j,n} \tag{8}$$

The  $1 \times N_f$  received signal at the  $k^{th}$  UE is given by

$$\mathbf{y}_k = [\mathbf{h}_{k,1}^H \mathbf{F}^{AB} \mathbf{F}_1^{DB} \mathbf{s}_1, \dots, \mathbf{h}_{k,N_f}^H \mathbf{F}^{AB} \mathbf{F}_{N_f}^{DB} \mathbf{s}_{N_f}] + \mathbf{w}_k, \tag{9}$$

Combining the signals for all UEs in a  $K$  dimensional received signal vector  $\mathbf{y} = [\mathbf{y}_1, \dots, \mathbf{y}_K]^H$ , we get the system equation as

$$\mathbf{Y} = \mathbf{H}^H \mathbf{F}^{AB} \mathbf{F}^{DB} \mathbf{S} + \mathbf{W}, \tag{10}$$

where  $\mathbf{Y}, \mathbf{W} \in \mathbb{C}^{K \times N_f}$ .

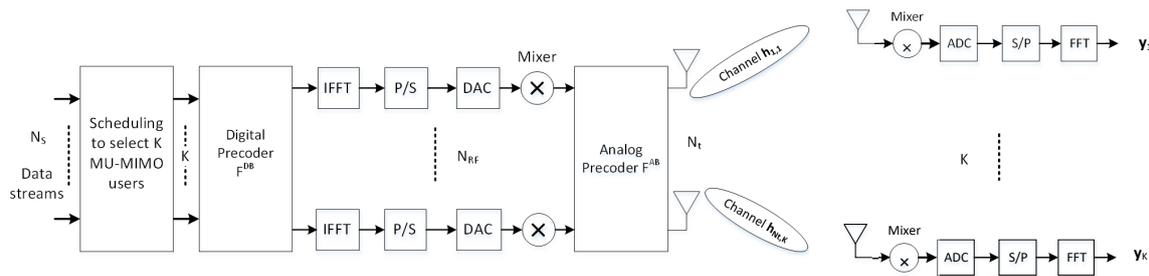


Figure 3. Block diagram of mmWave massive MIMO BS and UEs.

2.1. Channel Model

Generally, massive MIMO channel models are categorized in two types (i) analytical models and (ii) physical models [4]. Analytical models are commonly used for the theoretical analysis of wireless communication systems. The most commonly used analytical model is Kronecker channel model. It is a correlation-based model and characterizes the MIMO channel matrix in terms of the separate transmit and receive side spatial correlation matrices [24],

$$\begin{aligned} \mathbf{R}_H &= \mathbb{E}\{\text{vec}\{\mathbf{H}\}\text{vec}\{\mathbf{H}^H\}\} \\ &= \mathbf{R}_{tx} \otimes \mathbf{R}_{rx} \end{aligned} \tag{11}$$

under the above assumptions, the channel model  $\mathbf{H}$  is simplified to Kronecker model,

$$\mathbf{H} = \mathbf{R}_{rx}^{1/2} \mathbf{K} \mathbf{R}_{tx}^{1/2} \tag{12}$$

where  $\mathbf{K} \sim \mathcal{CN}(0,1)$  is an i.i.d. unit variance MIMO channel matrix,  $\mathbf{R}_{tx}$  and  $\mathbf{R}_{rx}$  are the transmit and receive correlation matrices, respectively. The transmit and receive correlation matrices are given as [24],

$$\mathbf{R}_{tx} = \mathbb{E}\{\mathbf{H}^H \mathbf{H}\}, \quad \mathbf{R}_{rx} = \mathbb{E}\{\mathbf{H} \mathbf{H}^H\} \tag{13}$$

The physical models explicitly model wave propagation parameters like the complex amplitude, DoD, DoA, and delay of an MPC [24,25]. MmWave propagation leads to limited spatial scattering due to the high free-space pathloss. In addition, the large tightly packed antenna arrays lead to high levels of antenna correlation. The sparse scattering and antennas spatial correlation makes many of the commonly used statistical fading distributions inaccurate for mmWave channel modeling. Therefore, we use extended Saleh-Valenzuela model, which accurately describes the mathematical structure present in mmWave channels [26,27]. For simplicity, we assume that each scattering cluster around the transmitter and receiver contributes a single propagation path [28].

In general, the mmWave MIMO channel matrix between the BS with  $N_t$  transmit antennas and a user  $k$  with  $n_r$  receive antennas in subchannel  $n$ , can be modeled as double directional channel,

$$\mathbf{H}_{k,n} = \sqrt{\frac{N_t n_r}{\rho_{k,n} L}} \sum_{l=0}^L \alpha_{k,n,l} \mathbf{a}(\phi_{k,n,l}) \mathbf{b}^H(\theta_{k,n,l}), \tag{14}$$

where  $L$  is the total number of multipaths,  $\alpha_{k,n,l}$  is the complex gain of the  $l^{th}$  path with i.i.d.  $\mathcal{CN}(0,1)$ , and  $\rho_{k,n}$  is the distance dependent pathloss between the BS and user  $k$  [29]. The LOS path is included with  $l = 0$ . Moreover,  $\mathbf{a}$  and  $\mathbf{b}$  are the receive and transmit steering vectors, respectively. The variables  $\phi_{k,n,l} \in [0, 2\pi)$  and  $\theta_{k,n,l} \in [0, 2\pi)$  are the  $l^{th}$  path's azimuth angles (boresight angles in the receive array and transmit array) of arrival and departure, respectively. The steering vectors are given by

$$\mathbf{a}(\phi_{k,n,l}) = \frac{1}{\sqrt{N_r}} [a_1(\phi_l), \dots, a_{n_r}(\phi_{k,n,l})]^T \tag{15}$$

$$\mathbf{b}(\theta_{k,n,l}) = \frac{1}{\sqrt{N_t}} [b_1(\theta_l), \dots, b_{N_t}(\theta_{k,n,l})]^T \tag{16}$$

The elements of transmit and receive steering vectors are given by

$$b_i(\theta_{k,n,l}) = e^{-j\omega\tau_{i,l}} = e^{-j2\pi(i-1)\frac{d_t}{\lambda} \sin(\theta_{k,n,l})}, \quad i = 1, 2, \dots, N_t \tag{17}$$

$$a_i(\phi_{k,n,l}) = e^{-j\omega\tau_{i,l}} = e^{-j2\pi(i-1)\frac{d_r}{\lambda} \sin(\phi_{k,n,l})}, \quad i = 1, 2, \dots, n_r \tag{18}$$

where  $\lambda$  is the wavelength,  $\omega = \frac{2\pi}{\lambda}$ ,  $\tau_i$  is the beamforming delay, and  $d_t$  and  $d_r$  are the antenna spacing at the transmitter and receiver, respectively.

The channel matrix in (14) can also be written in more compact form as

$$\mathbf{H}_{k,n} = \sqrt{\nu} \mathbf{A}_{k,n} \mathbf{D}_{k,n} \mathbf{B}_{k,n}^H \tag{19}$$

where  $\nu = \frac{N_t n_r}{\rho L}$  and  $\mathbf{A}_{k,n}$  and  $\mathbf{B}_{k,n}$  consist of stacked steering vectors of AoA and AoD, respectively, i.e.,  $\mathbf{A}_{k,n} = [\mathbf{a}(\phi_{k,n,1}), \mathbf{a}(\phi_{k,n,2}), \dots, \mathbf{a}(\phi_{k,n,L})]$  and  $\mathbf{B}_{k,n} = [\mathbf{b}(\theta_{k,n,1}), \mathbf{b}(\theta_{k,n,2}), \dots, \mathbf{b}(\theta_{k,n,L})]$ . The matrix  $\mathbf{D}_{k,n}$  is a diagonal matrix, given as  $\mathbf{D}_{k,n} = \text{diag}\{\alpha_{k,n,1}, \alpha_{k,n,2}, \dots, \alpha_{k,n,L}\}$ . The small scale fading at user  $k$  in subchannel  $n$  in multipath component (MPC) is given by  $\alpha_{k,n,l}$  with zero mean and variance  $\sigma_{k,n,l}^2$ . Assume that each MPC is i.i.d. such that  $\sum_{l=1}^L \sigma_{k,n,l}^2 = 1$ . We can express the channel model in (19) as

$$\mathbf{H}_{k,n} = \sqrt{\nu} \mathbf{A}_{k,n} \mathbf{\Sigma}_{k,n} \bar{\mathbf{D}}_{k,n} \mathbf{B}_{k,n}^H \tag{20}$$

where  $\mathbf{\Sigma}_{k,n} = \text{diag}\{\sigma_{k,n,1}^2, \sigma_{k,n,2}^2, \dots, \sigma_{k,n,L}^2\}$  and  $\bar{\mathbf{D}}_{k,n} = \text{diag}\{\bar{\alpha}_{k,n,1}, \bar{\alpha}_{k,n,2}, \dots, \bar{\alpha}_{k,n,L}\}$  with  $\bar{\alpha}_{k,n,l} = \frac{\alpha_{k,n,l}}{\sigma_{k,n,l}}$  such that  $E\{\bar{\alpha}_{k,n,l}\} = 0$  and  $E\{\bar{\alpha}_{k,n,l}^2\} = 1$ .

Substituting (20) in (13) and averaging over small scale fading, we get the transmit and receive correlation matrices for user  $k$  in the subchannel  $n$  as

$$\mathbf{R}_{tx,k,n} = \nu \mathbf{B}_{k,n} \mathbf{\Sigma}^2 \mathbf{B}_{k,n}^H \tag{21}$$

$$\mathbf{R}_{rx,k,n} = \nu \mathbf{A}_{k,n} \mathbf{\Sigma}^2 \mathbf{A}_{k,n}^H \tag{22}$$

For mmWave massive MIMO systems with large number of antennas, the steering vectors are asymptotically orthogonal to each other [6]:

$$\begin{aligned} \mathbf{a}^H(\phi_{k,n}) \mathbf{a}(\phi_{k,n'}) &\approx 0, \\ \mathbf{b}^H(\theta_{k,n}) \mathbf{b}(\theta_{k,n'}) &\approx 0. \end{aligned} \tag{23}$$

Moreover, in mmWave massive MIMO, acquisition of the instantaneous full CSI is not practical. Instead, an average CSI in terms of  $[\mathbf{A}_{k,n}]$ ,  $[\mathbf{B}_{k,n}]$ , and  $[\mathbf{\Sigma}_{k,n}]$  is a practical solution for the beamforming design because the coherence time of the channel statistics based CSI is of the order of few seconds or more as compared to the small scale of the order of milli-second [6].

### 2.2. Problem Formulation

The hybrid beamforming divides the beamforming matrix into two parts: covariance-based pre-beamforming matrix  $\mathbf{F}^{AB}$  realized by analog beamformers and the reduced dimension MU-MIMO digital precoding based on the effective channel  $\mathbf{H}^H \mathbf{F}^{AB}$  (omitting the subchannel subscript for simplicity). We assume that  $K$  users are divided into  $G$  groups, such that, the group  $g$  contains  $K_g$  number of users. Since users are near the ground level and surrounded by the scatterers compared to the scatterer-free elevated base-station, we assume one-ring model [1] and all users in group  $g$

experience the same azimuth center angle ( $\theta_g$ ) and angular spread ( $\Delta_g$ ). In this case,  $\mathbf{R}_{rx} = \mathbf{I}$  in (12), therefore, the channel covariance matrix of each user in group  $g$  is given by [30]

$$\mathbf{R}_g = \mathbb{E}\{\mathbf{h}_{gk} \mathbf{h}_{gk}^H\} \tag{24}$$

for which the eigenvalue decomposition gives

$$\mathbf{R}_g = \mathbf{U}_g \mathbf{\Lambda}_g \mathbf{U}_g^H \tag{25}$$

where  $\mathbf{U}_g \in \mathbb{C}^{N_t \times r_g}$  is a tall unitary matrix ( $\mathbf{U}_g \mathbf{U}_g^H = \mathbf{I}$ ) comprises the eigenvectors of  $\mathbf{R}_g$  and  $\mathbf{\Lambda}_g \in \mathbb{R}^{r_g \times r_g}$  is diagonal matrix with  $r_g$  nonzeros positive eigenvalues along the diagonal. The  $i, j$ -th element of covariance matrix  $\mathbf{R}_g$  represents the correlation between the channel coefficients antenna element  $i$  and  $j$  as

$$[\mathbf{R}_g]_{i,j} = \frac{1}{2\Delta_g} \int_{\theta_g - \Delta_g}^{\theta_g + \Delta_g} e^{-j2\pi \frac{d}{\lambda} (i-j) \sin \theta} d\theta, \tag{26}$$

where  $d$  is the distance between antenna elements of ULA and  $\lambda$  is the wavelength of carrier frequency. Using the Karhunen-Loeve representation, the channel vector of user  $k$  in group  $g$  is given as

$$\mathbf{h}_{gk} = \mathbf{U}_g \mathbf{\Lambda}_g^{1/2} \mathbf{z}_g = \mathbf{U}_g \tilde{\mathbf{h}}_{gk} \tag{27}$$

where  $\mathbf{z}_g \in \mathbb{C}^{N_t \times 1} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{r_g})$  and  $\tilde{\mathbf{h}}_{gk}$  is beam domain channel. For large  $N_t$ ,  $\mathbf{U}_g$  tends to discrete Fourier transform (DFT) matrix  $\Delta_{N_t} \in \mathbb{C}^{N_t \times N_t}$  [31]. Each column of  $\mathbf{U}_g$  represents one direction of angle-of-departure (AoD), i.e., a *beam*.

Alternatively, for the case, when dominant eigenvalues  $\hat{r}_g \leq r_g$ , then, the channel matrix can be written as ([13], Equation (5))

$$\mathbf{h}_{gk} = N_t^{1/2} \mathbf{R}_g^{1/2} \mathbf{z}_{gk} \tag{28}$$

The limited feedback-based hybrid beamforming consists of analog pre-beamforming matrix  $\mathbf{F}_g^{AB} \in \mathbb{C}^{N_t \times N_{RF,g}}$  responsible for spatial group formation and inter-group interference mitigation; and the digital multi-users precoding matrix  $\mathbf{F}_g^{DB} \in \mathbb{C}^{N_{RF,g} \times S_g}$  for spatial multiplexing inside the group and inter-user interference mitigation. Here,  $N_{RF,g}$  is the number of RF chains for group  $g$  such that  $S_g < N_{RF,g} < \hat{r}_g$  and  $S_g$  is the number of multi-carrier information symbols vectors for group  $g$  with  $N_{RF} = \sum_{g=1}^G N_{RF,g}$  and  $S = \sum_{g=1}^G S_g$ . The overall analog pre-beamforming matrix  $\mathbf{F}^{AB} \in \mathbb{C}^{N_t \times N_{RF}}$  is given by

$$\mathbf{F}^{AB} = [\mathbf{F}_1^{AB}, \dots, \mathbf{F}_G^{AB}], \tag{29}$$

and the overall digital beamforming matrix  $\mathbf{F}^{DB} \in \mathbb{C}^{N_{RF} \times N_s}$  is given by

$$\mathbf{F}^{DB} = \text{diag}[\mathbf{F}_1^{DB}, \dots, \mathbf{F}_G^{DB}] \tag{30}$$

and the overall channel matrix

$$\mathbf{H} = [\mathbf{H}_1, \dots, \mathbf{H}_G], \tag{31}$$

where the channel matrix of group  $g$  is defined as  $\mathbf{H}_g \triangleq [\mathbf{h}_{g1}, \dots, \mathbf{h}_{gK_g}]$ .

The analog pre-beamforming  $\mathbf{F}_g^{AB}$  is based on the slowly varying channel covariance matrix  $\mathbf{R}_g$  and can be implemented by the DFT matrix (when  $N_t$  is large), whereas, the digital beamformer  $\mathbf{F}_g^{DB}$  is based on the instantaneous channel information of the reduced dimension effective channel  $\mathbf{H}_g^H \mathbf{F}_g^{AB}$ . The overall effective channel is given by

$$\mathbf{H}_{eff}^H = \mathbf{H}^H \mathbf{F}^{AB} = \begin{bmatrix} \mathbf{H}_1^H \mathbf{F}_1^{AB} & \mathbf{H}_1^H \mathbf{F}_2^{AB} & \dots & \mathbf{H}_1^H \mathbf{F}_G^{AB} \\ \mathbf{H}_2^H \mathbf{F}_1^{AB} & \mathbf{H}_2^H \mathbf{F}_2^{AB} & \dots & \mathbf{H}_2^H \mathbf{F}_G^{AB} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_G^H \mathbf{F}_1^{AB} & \mathbf{H}_G^H \mathbf{F}_2^{AB} & \dots & \mathbf{H}_G^H \mathbf{F}_G^{AB} \end{bmatrix} \quad (32)$$

The excessive pilot transmission in downlink and feedback in uplink of FDD system can be reduced by only sending the group-wise average CSI based channel estimates in uplink. This is accomplished by using the diagonal elements  $\mathbf{H}_g^H \mathbf{F}_g^{AB}$  as feedback information with the size of  $K_g \times N_{RF,g}$  for  $g = 1, \dots, G$ . The analog pre-beamforming is designed in such a way that the other elements of matrix (32)  $\mathbf{H}_g^H \mathbf{F}_{g'}^{AB} \approx 0$  for all  $g' \neq g$ . This group-wise division creates virtual sectors, each group corresponds to a virtual sector [30].

The second order channel statistics-based RF beamformer  $\mathbf{F}^{AB}$  remains the same across multiple coherence blocks which gives the effective instantaneous channel between BS and user  $k$  as

$$\mathbf{h}_{n,gk,eff}^H \triangleq \mathbf{h}_{n,gk}^H \mathbf{F}_g^{AB}, \quad (33)$$

with  $\mathbf{h}_{n,gk,eff} \in \mathbb{C}^{N_{RF,g} \times 1}$ . Therefore, channel statistics-based CSI sufficiently reduces the feedback overhead on each user, otherwise, for instantaneous CSI, each user have to send the  $N_t \times 1$  size of channel estimate on the uplink channel. The covariance of effective channel  $\mathbf{h}_{n,gk,eff}^H$  is given by using (13) as,

$$\begin{aligned} \mathbb{E}\{\mathbf{h}_{n,gk,eff} \mathbf{h}_{n,gk,eff}^H\} &= \mathbb{E}\{\mathbf{F}_g^{ABH} \mathbf{h}_{n,gk} \mathbf{h}_{n,gk}^H \mathbf{F}_g^{AB}\} \\ &= \nu \mathbf{F}_g^{ABH} \mathbf{B}_{k,n} \Sigma^2 \mathbf{B}_{k,n}^H \mathbf{F}_g^{AB} \end{aligned} \quad (34)$$

The analog beamformer consists of columns of the DFT matrix, which can be easily implemented by phase shifter network. Therefore,  $\mathbf{F}_{n,g}^{AB}$  can be obtained by eigenvalue decomposition of channel covariance matrix. With the group-wise hybrid beamforming, the received signal  $\mathbf{y}_{g,n}$  for group  $g$  in subchannel  $n$  becomes

$$\mathbf{y}_{g,n} = \mathbf{H}_{g,n}^H \mathbf{F}_g^{AB} \mathbf{F}_{g,n}^{DB} \mathbf{s}_{g,n} + \mathbf{H}_{g,n}^H \sum_{g' \neq g} \mathbf{F}_{g'}^{AB} \mathbf{F}_{g',n}^{DB} \mathbf{s}_{g',n} + \mathbf{w}_{g,n} \quad (35)$$

and the received signal of user  $k$  in group  $g$  in subchannel  $n$  is given by

$$\begin{aligned} y_{gk,n} &= \mathbf{h}_{gk,n}^H \mathbf{F}_g^{AB} \mathbf{f}_{gk,n}^{DB} \mathbf{s}_{gk,n} + \underbrace{\mathbf{h}_{gk,n}^H \sum_{k' \neq k} \mathbf{F}_g^{AB} \mathbf{f}_{gk',n}^{DB} \mathbf{s}_{gk',n}}_{\text{Inter-user interference}} \\ &\quad + \underbrace{\mathbf{h}_{gk,n}^H \sum_{g' \neq g} \sum_j \mathbf{F}_{g'}^{AB} \mathbf{f}_{g',n}^{DB} \mathbf{s}_{g',n}}_{\text{Inter-group interference}} + w_{gk,n} \end{aligned} \quad (36)$$

The received signal to interference and noise ratio (SINR) at the user  $k$  in group  $g$  and subchannel  $n$  is given by

$$SINR_{gk,n} = \frac{|\mathbf{h}_{gk,n}^H \mathbf{F}_g^{AB} \mathbf{f}_{gk,n}^{DB}|^2}{|\mathbf{h}_{gk,n}^H|^2 \left( \sum_{k' \neq k} |\mathbf{F}_g^{AB} \mathbf{f}_{gk',n}^{DB}|^2 + \sum_{g' \neq g} \sum_j |\mathbf{F}_{g'}^{AB} \mathbf{f}_{g',n}^{DB}|^2 \right) + \sigma^2} \quad (37)$$

The spectral efficiency of user  $k$  in group  $g$  and subchannel  $n$  is expressed as

$$R_{gk,n} = \Psi_{gk,n} \log_2(1 + SINR_{gk,n}), \quad (38)$$

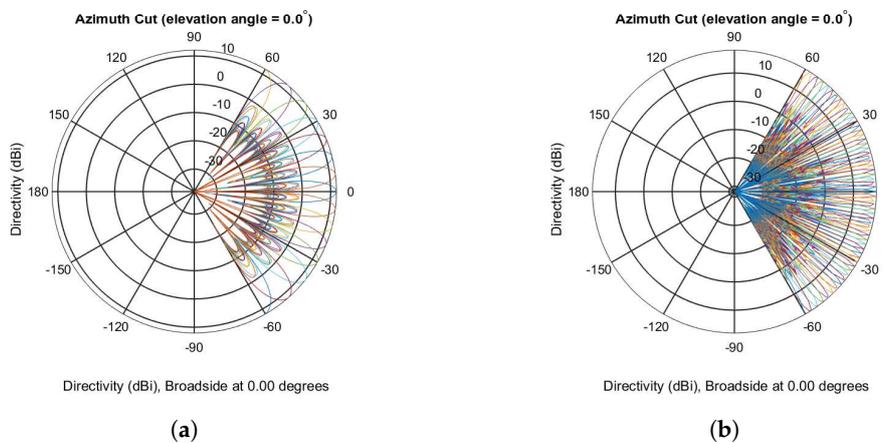
where  $\Psi_{gk,n}$  is the binary variable such that it is equal to 1 if user  $k$  is selected in group  $g$  in the subchannel  $n$ . In order to achieve balance tradeoff between throughput and fairness [32], we use proportional fairness (PF) based throughput maximization. We define per user proportional fairness metric as

$$U(\mathbf{F}_g^{AB}, \mathbf{f}_{gk,n}^{DB}) = \frac{R_{gk,n}(t)}{\bar{R}_{gk,n}(t)}, \forall n, g, k, \quad (39)$$

where  $\bar{R}_{gk,n}(t)$  is average throughput (moving average) over a past window of length  $T_w = 1/\alpha$  [33], as

$$\bar{R}_{gk,n}(t) = \alpha R_{gk,n}(t-1) + (1-\alpha)\bar{R}_{gk,n}(t-1), \quad (40)$$

The large number of antennas in massive MIMO systems enable the use of the eigenmodes of the channel covariance matrix, i.e.,  $\mathbf{B}_{k,n}$  comprises of the columns of the DFT matrix [6]. DFT-based beams with  $N_t = 16$  and  $N_t = 64$  are shown in Figure 4a,b, respectively.



**Figure 4.** DFT-based beams in a 120 sector. (a) DFT-based beams in a 120 sector with  $N_t = 16$ ; (b) DFT-based beams in a 120 sector with  $N_t = 64$ .

The beam steering matrix  $\mathbf{B}_{k,n}$  consists of selected columns of  $N_t \times N_t$  DFT matrix  $\Delta_{N_t}$  such that

$$\mathbf{F}_g^{AB} = \mathbf{B}_{k,n} = \Delta_{N_t} \mathbf{Y}_n \quad (41)$$

where  $\Delta_{N_t} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_{N_t}]$  consisting of all eigenmodes and  $\mathbf{Y}_n$  is an  $N_t \times r_R$  binary beam selection matrix, with  $r_R$  is the rank of the channel covariance matrix. The selection matrix  $\mathbf{Y}_n \in \mathbb{C}^{N_t \times r_R}$  with only a single one on each row and column such that  $\sum_i [\mathbf{Y}_n]_{i,j} = 1 \quad \forall j$ . Now we formulate our optimization problem for joint spatio-radio resource allocation and precoders design with the objective to maximize the utility function as

$$\begin{aligned} & \max_{\mathbf{F}_g^{AB}, \mathbf{f}_{gk,n}^{DB}} \sum_{g=1}^G \sum_{n=1}^{N_f} \sum_{k=1}^{K_g} U(\mathbf{F}_g^{AB}, \mathbf{f}_{gk,n}^{DB}) \quad (42) \\ & \text{subject to} \\ & \text{C1: } \text{tr}(\mathbf{F}_n^{DBH} \mathbf{F}_n^{ABH} \mathbf{F}_n^{AB} \mathbf{F}_n^{DB}) \leq P_n, \quad \forall n \\ & \text{C2: } \text{rank}(\mathbf{F}_n^{AB} \mathbf{F}_n^{DB}) \leq N_{RF}, \\ & \text{C3: } \sum_{k=1}^{K_g} \Psi_{k,n} \leq K_g, \quad \forall g, n, \\ & \text{C4: } \text{tr}(\mathbf{Y}_n) \leq G, \quad \forall n, \\ & \text{C5: } [\Psi_{k,n}], [\mathbf{Y}_n] \in \{0, 1\}. \end{aligned}$$

The above optimization problem is a mixed integer programming (MIP) problem with coupling between the digital and RF precoders in the power constraint. This MIP problem is NP-hard [14].

### 3. Relaxed-Convex Transformation

Though the above MIP optimization problem is NP-hard, it can be transformed to a relaxed convex optimization problem by (i) relaxing the binary integer constraints to real number between 0 and 1 [14], and (ii) decoupling the digital and analog precoders. For decoupling purpose, we make use of change of variables  $\mathbf{F}_n^{DB} = (\mathbf{F}^{ABH} \mathbf{F}^{AB})^{-\frac{1}{2}} \tilde{\mathbf{F}}_n^{DB}$ , where  $\tilde{\mathbf{F}}_n^{DB}$  is the *equivalent* digital precoder [34]. Thus, the problem in (42) can be written as

$$\begin{aligned} & \max_{\mathbf{F}_g^{AB}, \tilde{\mathbf{f}}_{gk,n}^{DB}} \sum_{g=1}^G \sum_{n=1}^{N_f} \sum_{k=1}^{K_g} \tilde{U}_{gk,n} \left( \mathbf{F}_g^{AB}, \mathbf{f}_{gk,n}^{DB} \right) \quad (43) \\ & \text{subject to} \\ & \text{C1 : } \text{tr}(\tilde{\mathbf{F}}_n^{DBH} \tilde{\mathbf{F}}_n^{DB}) \leq P_n, \quad \forall n \\ & \text{C2 : } \text{rank}(\tilde{\mathbf{F}}_n^{DB}) \leq N_{RF}, \\ & \text{C3 : } \sum_{k=1}^{K_g} \Psi_{k,n} \leq K_g, \quad \forall g, n, \\ & \text{C4 : } \text{tr}(\mathbf{Y}_n) \leq G, \quad \forall n, \\ & \text{C5 : } 0 \leq [\Psi_{k,n}] \leq 1, \quad 0 \leq [\mathbf{Y}_n] \leq 1, \end{aligned}$$

For a given RF precoder  $\mathbf{F}^{AB}$  and the knowledge of perfect CSI at the base-station, the digital precoder can be obtained by conventional MU-MIMO techniques, e.g., the *zero-forcing* and *block diagonalization* [15].

For the digital precoder, we adopt the ZF precoder for no multiuser interference among the users in each groups. The beamforming vector of user k is chosen to be orthogonal to the effective channel vectors of all the other users in the group. Zero-forcing is a suboptimal but low complexity approach within the linear precoders' class. ZF precoder is asymptotically optimal among all downlink beamforming techniques in high SNR region. It guarantees high spectral efficiency for large-scale antennas with low-complexity linear processing [35]. For  $N_t \gg N_r$ , it has shown that zero-forcing beamforming can achieve up to 98% of the non-linear dirty paper coding (DPC) capacity [36]. In order to make this paper self-contained, we describe the block diagonalization briefly. Since digital precoder is used to mitigate the multiuser interference within a groups and all groups are independent, we omit the subscript g. First we consider the downlink transmission over one subchannel n with the general case of BS with  $N_t$  antennas and  $K_n$  users with  $n_k$  antennas each, such that  $\sum_{k=1}^K n_k = N_r$ . The downlink channel on the subchannel n is expressed as  $N_r \times N_t$  matrix,

$$\mathbf{H}_{n,eff} = [\mathbf{H}_{1,n,eff}^T, \dots, \mathbf{H}_{K_n,n,eff}^T]^T \quad (44)$$

For user k, we define the following  $(N_r - n_k) \times N_t$  channel matrix

$$\mathbf{H}'_{k,n,eff} = [\mathbf{H}_{1,n,eff}^T, \dots, \mathbf{H}_{K-1,n,eff}^T, \mathbf{H}_{k+1,n,eff}^T, \dots, \mathbf{H}_{K_n,n,eff}^T]^T \quad (45)$$

Let the rank of  $\mathbf{H}'_{k,n,eff}$  be denoted by  $r'_{k,n}$ , then the nullspace of  $\mathbf{H}'_{k,n,eff}$  has dimension  $N_t - r'_{k,n} \geq n_k$ . Performing the SVD of each user's channel matrix in subchannel n leads to the following

$$\mathbf{H}'_{k,n,eff} = \mathbf{U}'_{k,n} \Sigma'_{k,n} \mathbf{V}'_{k,n}{}^H = \mathbf{U}'_{k,n} \Sigma'_{k,n} [\mathbf{V}'_{k,n}{}^{(1)} \mathbf{V}'_{k,n}{}^{(0)}]^H, \quad (46)$$

where  $\mathbf{U}'_{k,n}$  and  $\mathbf{V}'_{k,n}$  are the unitary matrices. The columns of  $\mathbf{U}'_{k,n}$  are the left singular vectors of  $\mathbf{H}'_{k,n,eff}$ , the columns of  $\mathbf{V}'_{k,n}$  are the right singular vectors of  $\mathbf{H}'_{k,n,eff}$ , and  $\Sigma'_{k,n}$  is a diagonal matrix in

which the diagonal entries are the singular values of  $\mathbf{H}'_{k,n,eff}$ . In the last equality of (46),  $\mathbf{V}'_{k,n}(1)$  holds the first  $r'_{k,n}$  right singular vectors of  $\mathbf{H}'_{k,n,eff}$  and  $\mathbf{V}'_{k,n}(0)$  contains the  $N_t - r'_{k,n}$  singular vectors of  $\mathbf{H}'_{k,n,eff}$  which are in the nullspace of  $\mathbf{H}'_{k,n,eff}$ . The columns of  $\mathbf{V}'_{k,n}(0)$  are best suited for user  $k$  beamforming matrix  $\mathbf{F}_{k,n}^{DB}$ , because they will provide zero interference at other UEs. Usually  $\mathbf{V}'_{k,n}(0)$  contains more number of columns than the  $n_k$ , therefore we use some linear combinations of the columns of  $\mathbf{V}'_{k,n}(0)$  to make at most  $n_k$  columns.

$$\mathbf{H}_{k,n,eff} \mathbf{V}'_{k,n}(0) = \mathbf{U}_{k,n} \begin{bmatrix} \boldsymbol{\Sigma}_{k,n} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_{k,n}(1) & \mathbf{V}_{k,n}(0) \end{bmatrix}^H, \tag{47}$$

where  $\mathbf{H}_{k,n,eff} \mathbf{V}'_{k,n}(0)$  gives the matrix with columns as the linear combinations of the columns of  $\mathbf{V}'_{k,n}(0)$ . The right hand side of the equation is the SVD of  $\mathbf{H}_{k,n,eff} \mathbf{V}'_{k,n}(0)$ , where  $\boldsymbol{\Sigma}_{k,n}$  is  $r_{k,n} \times r_{k,n}$  diagonal matrix and  $\mathbf{V}_{k,n}(1)$  represents the  $r_{k,n}$  singular vectors with nonzero singular values of  $\mathbf{H}_{k,n,eff} \mathbf{V}'_{k,n}(0)$ . The Equation (47) can also be written as,

$$\begin{aligned} \mathbf{H}_{k,n,eff} \mathbf{V}'_{k,n}(0) &= \mathbf{U}_{k,n} \boldsymbol{\Sigma}_{k,n} \mathbf{V}_{k,n}(1)^H \\ \mathbf{H}_{k,n,eff} &= \mathbf{U}_{k,n} \boldsymbol{\Sigma}_{k,n} \mathbf{V}_{k,n}(1)^H \mathbf{V}'_{k,n}(0)^H \end{aligned} \tag{48}$$

The transmit beamforming matrix that maximizes the user  $k$  throughput without any inter-user interference is obtained as,

$$\mathbf{F}_{k,n}^{DB} = \mathbf{V}'_{k,n}(0) \mathbf{V}_{k,n}(1) \tag{49}$$

The transmit digital beamforming matrix for subchannel  $n$  is defined as

$$\mathbf{F}_n^{DB} = [\mathbf{F}_{1,n}^{DB}, \dots, \mathbf{F}_{K,n}^{DB}] \mathbf{P}_n^{1/2}, \tag{50}$$

where  $\mathbf{F}_{k,n}^B \mathbf{F}_{k,n}^B{}^H = \mathbf{I}$ ,  $1 \leq k \leq K_n$  and  $\mathbf{P}_n$  is a block diagonal matrix whose elements scale the power allocated to each interference-free virtual subchannel for all UEs. The receive combining matrix for this user is  $\mathbf{U}_{k,n}$  [37].

In the case of single antenna users, complete diagonalization is achieved entirely at the BS by channel inversion, i.e.,  $\mathbf{F}_n^{DB} = (\mathbf{H}_{n,eff}^H)^{\dagger}$ , where  $(\mathbf{H}_{n,eff}^H)^{\dagger}$  is the pseudo-inverse of  $\mathbf{H}_{n,eff}^H$  [38].

$$\begin{aligned} \mathbf{F}_n^{DB} &= \beta_n \left( \mathbf{H}_{n,eff}^H \right)^{\dagger} \\ [\mathbf{f}_{1,n}^{DB}, \dots, \mathbf{f}_{K,n}^{DB}] &= \beta_n \left[ [\mathbf{h}_{1,n}, \dots, \mathbf{h}_{K,n}]^H \mathbf{F}^{AB} \right]^{\dagger} \end{aligned} \tag{51}$$

where  $\beta_n$  is a normalization factor chosen to satisfy the power constraint and is given by

$$\beta_n^2 = \frac{K}{\|\mathbf{F}^{AB} \mathbf{F}_n^{DB}\|_F^2} \tag{52}$$

Using the definition of the pseudo-inverse, we get,

$$\mathbf{F}_n^{DB} = \beta_n \left( \mathbf{H}_{n,eff} \mathbf{H}_{n,eff}^H + N_{RFQ} \mathbf{I}_{N_{RF}} \right)^{-1} \mathbf{H}_{n,eff} \tag{53}$$

where  $\rho$  is the regularization parameter,  $\rho = 0$  for ZF precoding and  $\rho = \frac{N_s}{N_{RF}\eta}$  for regularized ZF, with  $\eta = P_{T,n}/\sigma^2$ . Lastly, introducing the group subscript again, the SINR of user  $g_k$  is given by

$$SINR_{g_k,n} = \frac{\frac{P_{T,n}}{K_g} \beta^2 |\mathbf{h}_{g_k,n}^H \mathbf{F}_g^{AB} \mathbf{F}_{g,n}^{DB} \mathbf{F}_{g,n}^{DBH} \mathbf{F}_g^{ABH} \mathbf{h}_{g_k,n}|^2}{|\mathbf{h}_{g_k,n}^H|^2 \left( \sum_{k' \neq k} |\mathbf{F}_g^{AB} \mathbf{f}_{g_k',n}^{DB}|^2 + \sum_{g' \neq g} \sum_j |\mathbf{F}_{g'}^{AB} \mathbf{f}_{g',n}^{DB}|^2 \right) + \sigma^2} \quad (54)$$

and the PF sum rate is calculated as

$$\hat{U} = \sum_{g=1}^G \sum_{n=1}^{N_f} \sum_{k=1}^{K_g} \tilde{U}_{g_k,n}. \quad (55)$$

#### 4. Suboptimal Solution

Joint optimization of analog and digital beamformers is challenging because they use different channel information for the design of analog and digital beamformers. Hybrid beamforming methods consider decoupled designs of analog and digital beamforming to reduce the complexity of joint optimization, but the main challenge remains the use of different channel information. To approximate the optimal solution to this mixed integer programming problem, we summarize our proposed algorithm below:

The analog precoder is formed by selecting  $K_g$  columns of DFT matrix of eigenvectors of channel covariance matrix  $\mathbf{R}_g$  of users' group  $g$  in (41) to minimize the inter-group interference  $\mathbf{I}_g$ ,

$$\begin{aligned} \min_{\mathbf{Y}} \quad & \mathbf{I}_g(\mathbf{Y}) \\ \text{subject to} \quad & \\ & \text{tr}(\mathbf{Y}_n) \leq G, \quad \forall n, \end{aligned} \quad (56)$$

where  $\mathbf{I}_g = \sum_{i=1}^G \sum_{j=1, j \neq i}^G |\mathbf{H}_i^H \mathbf{F}_j^{AB}|^2$ . To solve the MIP problem, we divide the solution into two parts.

In the first part, we get the analog precoder using the selected columns of the DFT matrix which maximize the PF sum-rate. The inherent benefit of the DFT matrix is its constant modulus which enables the use of analog phase shifters and RF switches to realize the analog beamforming. In the second part, for the given analog precoder, intra-group users scheduling is performed and a ZF digital precoder is designed to maximize the sum-rate utility function. The decoupling of the analog and digital precoders design makes the solution suboptimal but tractable [34]. The joint hybrid beamforming and user scheduling Algorithm 1 takes  $K_g, N_f, N_{RF}, N_t$ , and  $K$ . It generates the analog beamforming matrix  $\mathbf{F}^{AB}$ , digital beamforming matrix  $\mathbf{F}^{DB}$ , and the set of users in each group. The first part of the algorithm (line 9 to line 19) forms DFT/eigenmodes-based analog beamforming using limited statistical CSI feedback from the users. Beam and users pairing within each group is taken place in this part of the algorithm. The *while* loop at line 12 executes till all the binary combinations in  $N_t \times N_t$  are exhausted with the condition that each column contain exactly one binary 1 and total number of 1s are equal to the number of streams (or number of RF chains). The second part (starts from line 39) assigns the radio resources to users to maximize the utility function.

**Algorithm 1** Joint Resource Allocation and Hybrid Beamforming Design Algorithm.

---

```

1: Inputs
2:  $K_g$ : Number of users per group
3:  $N_f$ : Number of subchannels
4:  $K_n$ : Number of UEs to be scheduled in subchannel  $n$ 
5:  $N_{RF}$ : Number of RF chains
6:  $N_t$ : Number of transmit antennas
7: Initialization
8:  $\mathcal{K}_g = \{1, \dots, K_g\}$ ,  $U(\mathbf{F}_n^{DB})^{last} = 0$ ,  $\mathbf{Y}_g^{last} = \mathbf{0}$ ,  $\mathcal{K}_n = \{\}$ ,  $\forall n$ 
   {Average CSI-based Beam Selection and RF Precoder Design}
9: Evaluate channel covariance matrix  $\mathbf{R}_g$  from CSI at users:  $\mathbf{R}_g = \mathbb{E}\{\mathbf{h}_{gk} \mathbf{h}_{gk}^H\}$ 
10: Find transmit steering  $\mathbf{B}_{\mathcal{K}_n}$  by Eigenvector based unitary matrix  $\mathbf{U}_g$  of  $\mathbf{R}_g$  or DFT matrix  $\Delta_{N_t}$ .
11: while  $g \leq G$  do
12:   while  $tr(\mathbf{Y}_g) \leq N_{RF,g}$ , set  $\{[Y_g]_{i,j}, \dots, [Y_g]_{N_t, N_t}\} = \{0, 1\}$  do
13:     Calculate  $U(\mathbf{Y}_g)$ , employing zero forcing  $\mathbf{F}_{n,g}^{DB}$ 
14:     if  $U(\mathbf{Y}_g) \geq U(\mathbf{Y}_g^{last})$  then
15:        $\mathbf{Y}_g^{last} \leftarrow \mathbf{Y}_g$ 
16:     end if
17:   end while
18: end while
19: Get  $\mathbf{F}_g^{AB} = \Delta_{N_t} \mathbf{Y}_g^{last}$ 
   {Instantaneous CSI-based Users Selection and Digital Precoder Design}
20: while  $n \leq N_f$  do
21:   while  $b \leq N_{RF,g}$  do
22:     while  $k \leq K_g$  do
23:       Compute  $U(\mathbf{F}_n^{DB}) = U(\mathbf{F}_n^{DB}) + U(\mathbf{f}_{k',n}^{DB})$ ,  $\forall k' \in \mathcal{K}_g$ 
24:        $k^* = \arg \max_k \{U(\mathbf{F}_n^{DB})\}$ 
25:        $\Psi_{k^*,n} = 1$ 
26:       Update  $\mathbf{F}_n^{DB}$  and  $U(\mathbf{F}_n^{DB})^{update}$  with user  $k^*$ 
27:       if  $U(\mathbf{F}_n^{DB})^{updated} \geq U(\mathbf{F}_n^{DB})^{last}$  then
28:          $\mathcal{K}_n = \mathcal{K}_n \cup \{k^*\}$ 
29:          $\mathcal{K}_g = \mathcal{K}_g - \{k^*\}$ 
30:          $U(\mathbf{F}_n^{DB})^{last} = U(\mathbf{F}_n^{DB})^{updated}$ 
31:          $k++$ 
32:       else
33:         break
34:       end if
35:     end while
36:      $b++$ 
37:   end while
38:    $n++$ 
39: end while
40: Stack the beamforming matrices  $\mathbf{F}^{DB} = [\mathbf{F}_1^{DB}, \dots, \mathbf{F}_{N_f}^{DB}]$ 
41: Output
42:  $\mathbf{F}^{AB} = [\mathbf{F}_1^{AB}, \dots, \mathbf{F}_G^{AB}]$ 
43:  $\mathbf{F}^{DB} = [\mathbf{F}_1^{DB}, \dots, \mathbf{F}_{N_f}^{DB}]$ 
44:  $[\mathcal{K}_1, \dots, \mathcal{K}_{N_f}]$ 

```

---

The Algorithm 1 is illustrated in flowchart Figure 5.

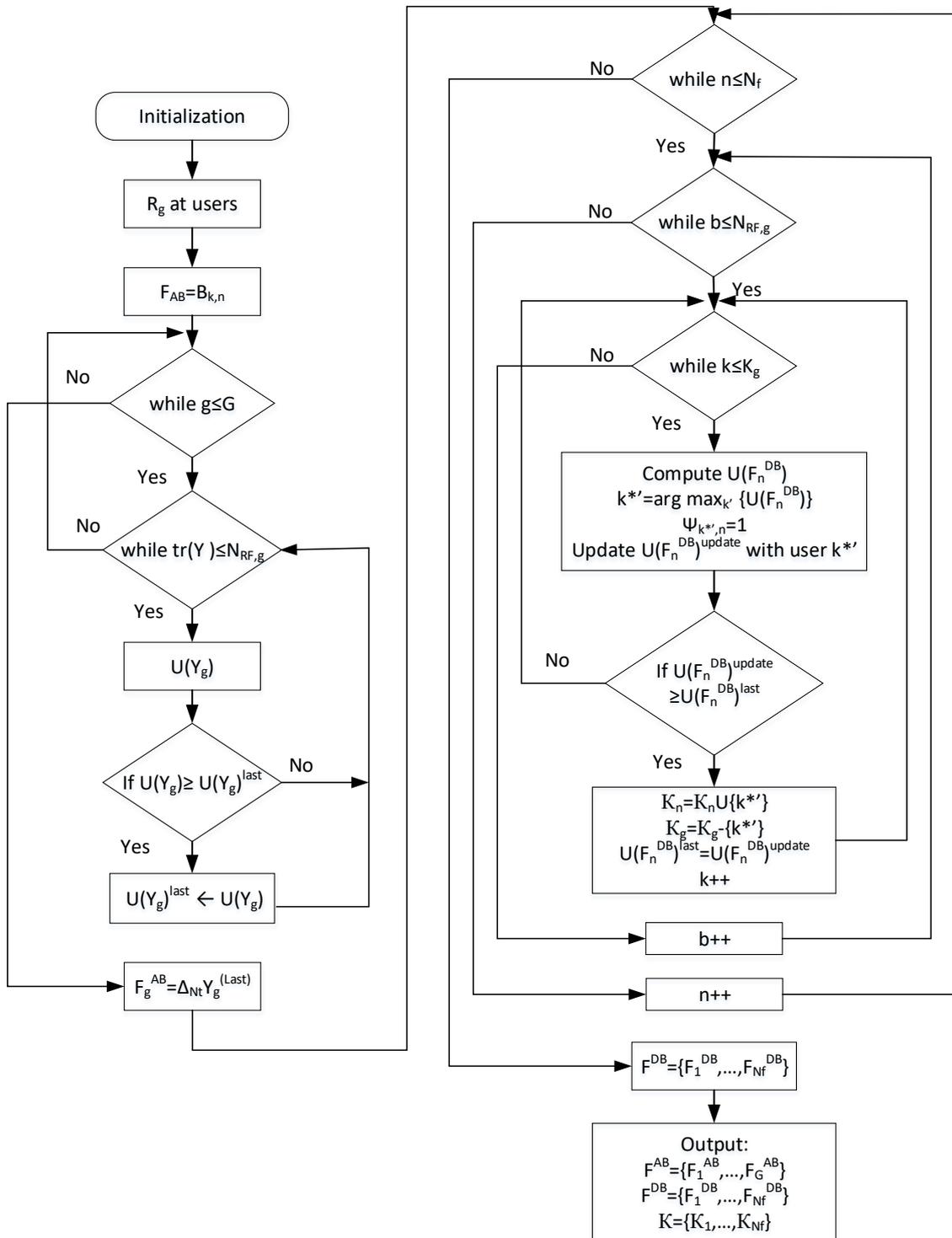


Figure 5. Flowchart illustration of Algorithm 1.

### 5. Machine Learning: K-Means Based Optimal Users Grouping for Analog Beamforming

In this section, we use machine learning technique to group the users. Then, the DFT based fixed switched-beams are used to realize the analog beamforming matrix. The joint users scheduling and hybrid beamforming architecture with ML-based users grouping is shown in Figure 6.

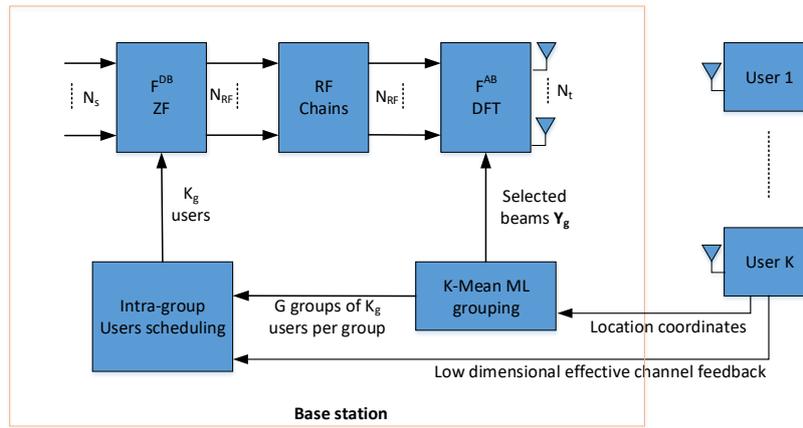


Figure 6. Joint users scheduling and hybrid beamforming architecture with ML-based users grouping.

Machine learning algorithms can broadly be divided into two main categories, namely supervised learning and unsupervised learning algorithms. The former class of algorithms learn by training on the input labeled examples, called training dataset,  $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), (x^{(3)}, y^{(3)}), \dots, (x^{(m)}, y^{(m)})\}$ , where the  $i^{th}$  example  $(x^{(i)}, y^{(i)})$  consists of the  $i^{th}$  instance of feature vector  $x^{(i)}$  and the corresponding label  $y^{(i)}$ . Given a labeled training dataset, these algorithms try to find the decision boundary that separates the positive and negative labeled examples by fitting a hypothesis to the input dataset. Unsupervised machine learning algorithms, on the other hand, are given an unlabeled input dataset. These algorithms are used for extracting information or features from the dataset. These features might be related, but not confined, to the underlying structures or patterns in the input data, relationships in data items, grouping/clustering of data items, etc. Discovered features are meant to provide a deeper insight into the input dataset that can subsequently be exploited for achieving specific goals. Clustering algorithms make an important part of unsupervised learning where the input examples are grouped into two or more separate clusters based on some features. The K-Means (KM) algorithm, is probably the most popular clustering algorithm. It is an iterative algorithm that starts with a set of initial centroids given to it as input. During each iteration, it performs the following two steps.

1. **Assign Cluster:** For every user, the algorithm computes the distance between the user and every centroid. The user is then associated to the cluster with the closest centroid. During this step, a user might change its association from one cluster to another one.
2. **Recompute centroids:** Once all users have been associated to their respective cluster, the new position of centroid for every cluster is then calculated.

Figure 7a depicts how the cluster centroids keep moving across iterations until the system stabilizes for an example network consisting of thirty users being grouped in five clusters. The system becomes stable in only five iterations and the final cluster layout is shown in Figure 7b.

Let us define the following notations to be used later in this section.

- $K$  = Total number of clusters being formed.
- $x^{(i)}$  = Location coordinates of user  $u^{(i)}$ . In our case,  $x^{(i)} \in \mathbb{R}^2$
- $c^{(i)}$  = Cluster to which the user  $u^{(i)}$  is currently associated.
- $\mu_k$  = Centroid of  $k^{th}$  cluster,  $\mu_k \in \mathbb{R}^2$
- $\mu_{c^{(i)}}$  = Centroid of the cluster to which the user  $u^{(i)}$  is currently associated.

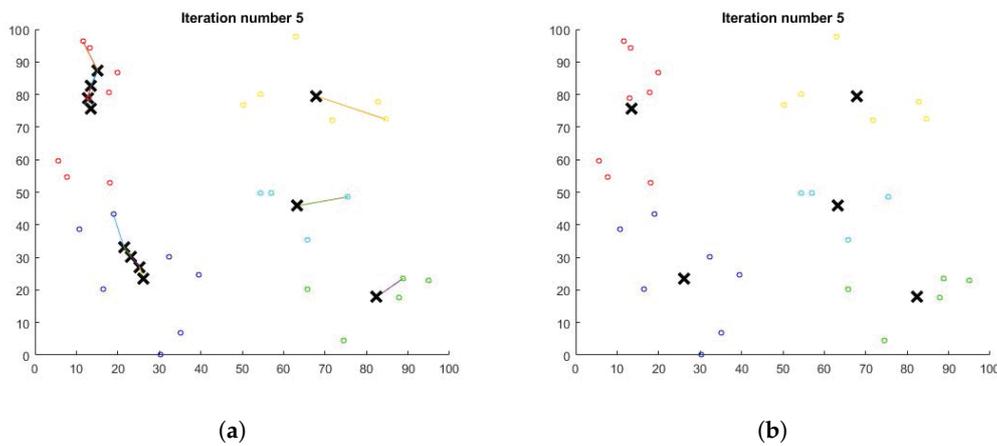
Now the cost function  $J$  can be defined as

$$J(c^{(1)}, c^{(2)}, \dots, c^{(m)}, \mu_1, \mu_2, \dots, \mu_K) = \frac{1}{m} \sum_{i=1}^m \|x^{(i)} - u_{(c^i)}\|^2 \tag{57}$$

with the following optimization objective function.

$$\min_{c^{(1)}, \dots, c^{(m)}, \mu_1, \dots, \mu_K} J(c^{(1)}, c^{(2)}, \dots, c^{(m)}, \mu_1, \mu_2, \dots, \mu_K)$$

It may be pointed out that Equation (57) allows us to compare multiple clustering layouts based on their cost and select the one with the lowest cost. The above optimization objective function constitutes a non-convex and NP-hard problem because it has many possible local minima and integer optimization variable  $c^{(i)}$ . The KM algorithm heuristically optimize this function by alternate minimization method. It iterates between two steps (Assign cluster and Recompute centroids) as described above.



**Figure 7.** Change in position of centroids as K-Means clustering algorithm progresses. (a) shows the transition of cluster centroids (shown as crosses) up to iteration 5, whereas, (b) shows only the final stable state after iteration 5. In the figures, cross-sign represents cluster centroids and the colored-circle-sign represents the user associated with the same group or cluster.

In this section, we use the KM algorithm for optimal clustering of  $m$  users competing for resources in a particular cell. The clustering is performed based on their geographic location, thus our input dataset  $\{u^{(1)}, u^{(2)}, u^{(3)}, \dots, u^{(m)}\}$  has  $m$  vectors  $u^{(i)}, 1 \leq i \leq m$ , consisting of location coordinates, of  $i$ th user. For the sake of simplicity, we assume these users are deployed in a two dimensional area, i.e., a plane and so  $u^{(i)} = (x_1^{(i)}, x_2^{(i)})$ , i.e., an ordered pair of location coordinates. Our clustering algorithm is summarized in Algorithm 2.

The proposed algorithm takes the location coordinates of  $m$  users as input. It also takes two numbers  $min_k$  and  $max_k$  as additional input. The algorithm outputs the best number of clusters,  $k$ , such that  $min_k \leq k \leq max_k$ , and corresponding members of each cluster. It starts with  $k = min_k$  and randomly selects  $k$  user locations as the initial centroids (line 6). It assigns the closest centroid to each user (line 8) and then computes new centroids by calculating the center/average location of all nodes in each cluster (line 11). So, in effect, the location of centroids keeps moving in successive iterations. It repeats the above two steps until the change in centroid positions is zero or negligible. We repeat the test  $max_t$  times with a new set of randomly chosen initial centroids every time. During every test, the discovered centroids, corresponding centroid assignment to users, and the cost are saved (lines 14–16) for later comparison. After running the loop for  $max_t$  times, we select and store the best  $k$  centroids resulting from the test with the lowest cost while discarding the remaining (lines 19–21).

The same is repeated for the next value of  $k$ , i.e.,  $k = k + 1$ , until  $k > \max_k$ . At the end we have  $\text{cnt} = \max_k - \max_k$  vectors  $\mu_k$ , one for each value of  $k$ , the corresponding assignment vector  $\mathbf{a}_k$  and cost  $c_k$ . Finally, we choose the vector  $\mu$  having the lowest cost and corresponding assignment vector  $\mathbf{a}$  among  $\text{cnt}$  stored cases. That is the best number of clusters and corresponding centroids that the algorithm found.

---

**Algorithm 2** K-Means based users grouping algorithm.

---

```

1:  $\text{cnt} = 0$ 
2: for  $k = \min_k \max_k$  do
3:    $\text{cnt} = \text{cnt} + 1$ 
4:   for  $t = 1 : \max_t$  do
5:     repeat
6:       Randomly choose initial  $k$  centroids  $\mu_1, \mu_2, \mu_3, \dots, \mu_k$ 
7:       for  $i = 1 : m$  do
8:          $a^{(i)} = j, 1 \leq j \leq k$ , such that  $\mu_j$  is the centroid closest to  $u^{(i)}$ 
9:       end for
10:      for  $l = 1 : k$  do
11:         $\mu_l = \text{mean of all users/points } u^{(i)} \text{ assigned to } l\text{th centroid}$ 
12:      end for
13:      until converges
14:       $\mu^{(t)} = (\mu_1, \mu_2, \mu_3, \dots, \mu_k)$ 
15:       $\mathbf{a}^{(t)} = (a^{(1)}, a^{(2)}, a^{(3)}, \dots, a^{(m)})$ 
16:       $c^{(t)} = \text{cost}(\mu_1, \mu_2, \mu_3, \dots, \mu_k)$ 
17:    end for
18:     $\text{idx} = \arg_{\min} \{c^{(t)}, 1 \leq t \leq \max_t\}$ 
19:     $\mu_k^{(k)} = \mu^{(\text{idx})}, 1 \leq \text{idx} \leq \max_t$ 
20:     $\mathbf{a}_k^{(k)} = \mathbf{a}^{(\text{idx})}, 1 \leq \text{idx} \leq \max_t$ 
21:     $c_k^{(k)} = c^{(\text{idx})}, 1 \leq \text{idx} \leq \max_t$ 
22:  end for
23:   $\text{index} = \arg_{\min} \{c_k^{(k)}, 1 \leq k \leq \text{cnt}\}$ 
24:   $\mu = \mu_k^{(\text{index})}, 1 \leq \text{index} \leq \text{cnt}$ 
25:   $\mathbf{a} = \mathbf{a}_k^{(\text{index})}, 1 \leq \text{index} \leq \text{cnt}$ 
26:   $n = \text{index}$ 

```

---

After the groups formation, BS sends this information to all users, where users use this information to form reduced average statistical CSI. For example, a user in a group of 5, needs to send the average statistical CSI only after 1/5 of regular feedback interval time.

## 6. Simulation Results

Consider the downlink of a multiuser massive MIMO single cell with three 120 degree sectors. We neglect inter-sector interference and focus on a single 120 degree sector served by a ULA of  $N_t = 64$  isotropic antennas at BS. The users grouping forms virtual sectors inside 120 physical sector.

In simulation, the results are obtained by averaging over 100 drops. In each drop we randomly generate spatial correlation matrices  $\mathbf{R}_g$ . For each realization of spatial correlation matrix  $\mathbf{R}_g$ , we simulate 1000 realizations of instantaneous channel  $\mathbf{H}$ .

The joint spatio-radio scheduling and hybrid precoder scheme first forms the users groups and then selects the beams that maximizes the sum-rate through downlink training process. Secondly, it calculates the ZF based digital precoder using low dimensional effective channel feedback from the users.

Figure 8 shows the CDF of the non-zero eigenvalues of channel covariance matrix. Notice that approximately 50% of the non-zero eigenvalues are close to zero. The sum-rate increases as the number of groups increases at the cost of increased feedback overhead as shown in Figure 9. Using machine learning technique in Section 5 we can get optimal number of groups from channel covariance feedback. This results in increased sum-rate with substantial reduced feedback. The optimal  $G = 3$  gives 27.6% increase in sum-rate compared to when  $G = 1$  and 62.5% decrease in feedback overhead compared to  $G = 8$ . The comparison of performance of ML-based users grouping with previous work cannot be provided because there is no previous work that uses ML-based technique to reduce the CSI feedback overhead in massive MIMO systems. Many papers use users grouping in massive MIMO hybrid beamforming [3,5,39,40], but they do not utilize ML-based users grouping. Therefore, we have compared our proposed solution with two benchmarks of full-CSI ( $G = K$ ) and coarse-CSI ( $G = 1$ ). Figure 10 shows sum-rate with number of users at 10dB SNR. For a fixed number of groups  $G = 3$ , the increase in number of users, increases number of users per group. Due to the fixed number of groups, the feedback overhead remains the constant. Sum-rate is increasing with users because we assumed  $N_{RF} = N_s = K$ . If we fix the number of RF chains to some hardware limit, then the sum-rate will saturate at specific number of users. It can be seen in Figure 10, that increasing number of users per group decreases the slop of the sum-rate for limited CSI schemes. This decrease is due the increase in intra-group interference.

Sum-rate also depends on the number of RF chains but this dependence is not linear as shown in Figure 11. This figure shows sum-rate variation with number of RF chains  $N_{RF}$  when  $N_s = 8$ ,  $K = 8$ ,  $N_t = 64$ , and  $SNR = 10$  dB. Sum rate increases with number of RF chains because it yields better conditioned effective channel matrix. It can be seen that the spectral efficiency does not increase monotonically with  $N_{RF}$  and saturates at  $N_{RF} = N_t$  where hybrid precoding is turned to the pure digital precoding. The increase in spectral efficiency with the number of RF chains comes at the cost of higher dimensional effective channel feedback overhead and power consumption in RF chains.

The spectral efficiency of the proposed scheme also varies with number of transmit antennas as shown in Figure 12. In the figure,  $N_{RF} = N_s = K = 8$ ,  $SNR = 10$  dB, and BS has 16, 64, 128 or 256 ULA antennas. The performance gain increases with the increase in number of transmit antennas because large antennas array increases the resolution of the transmit beams (also depicts in figure 4) and, hence, decreases the potential of inter-beams interference.

In general, the spectral efficiency is a function of SNR and for the  $SNR = 10$  dB, our ML-based users grouping and hybrid beamforming scheme gives 27.6% increased sum-rate at the cost of 33.3% extra feedback overhead as compared to the coarse-CSI case ( $G = 1$ ). Our proposed scheme incurs 62.5% reduced feedback at the cost of 25.2% reduction in sum-rate as compared to the full-CSI case ( $G = K$ ).

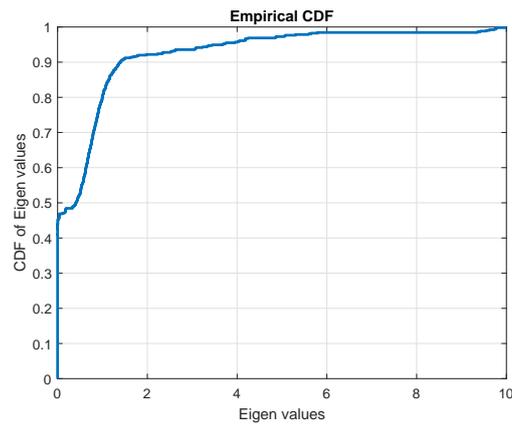


Figure 8. CDF of non-zero eigenvalues of channel covariance matrix  $\mathbf{R}_g$  for  $N_t = 64$ .

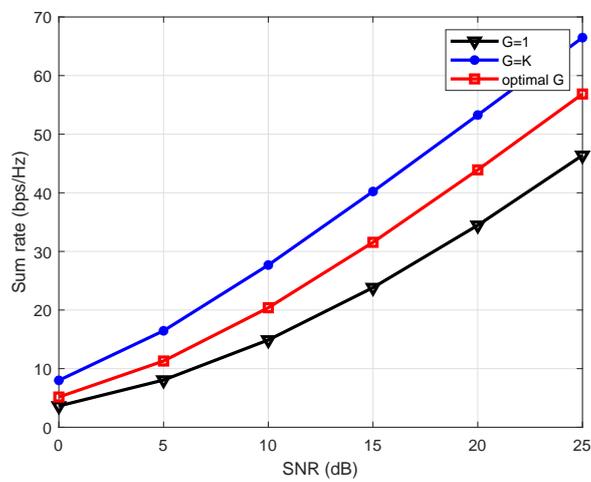


Figure 9. Sum-rate Vs SNR with different number of groups,  $K = 8$ .

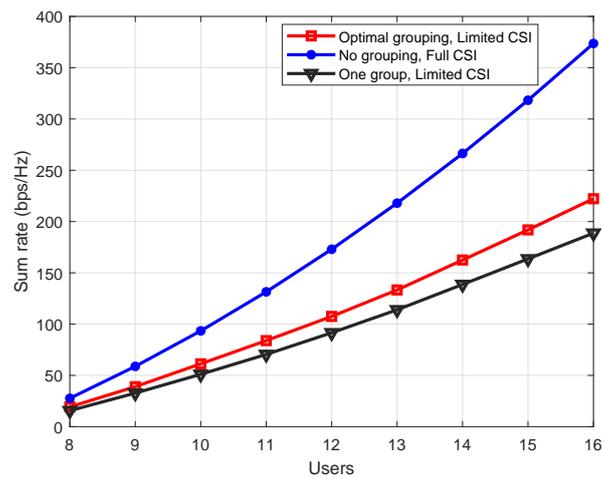
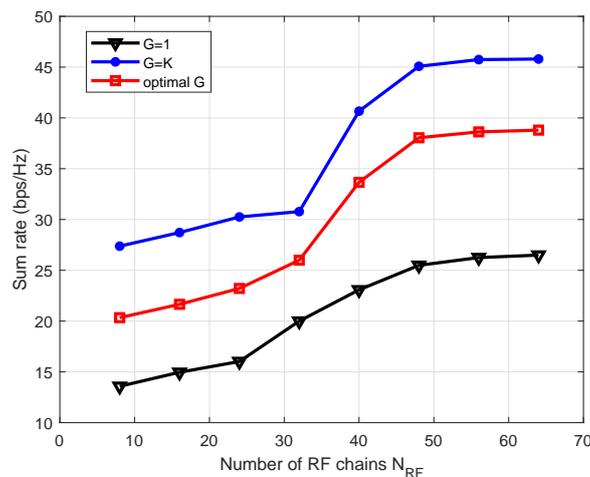
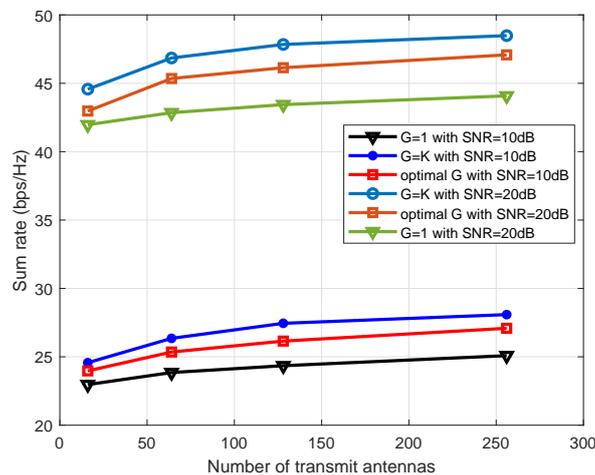


Figure 10. Sum-rate Vs number of users with different number of groups and CSI,  $SNR = 10$  dB.



**Figure 11.** Sum-rate Vs number of RF chains. Number of users  $K = 8$ , number of transmit antennas  $N_t = 64$ , and  $SNR = 10$  dB.



**Figure 12.** Sum-rate Vs number of transmit antennas. The number of RF chains at BS is fixed at 8, number of users is 8 and the results are obtained for  $SNR = 10$  dB and  $SNR = 20$  dB.

## 7. Conclusions

This paper studied the limited feedback two-stage hybrid beamforming for decomposing the precoding matrix at the base-station. The huge channel state information feedback of massive MIMO has been reduced by the channel covariance-based RF precoding and beam selection. The well-known regularized block diagonalization can mitigate the inter-group interference, but requires substantial feedback. We used K-mean algorithm based unsupervised machine learning technique for users grouping and channel covariance-based eigenmodes/discrete Fourier transforms to reduce the feedback overhead and designed a simplified analog precoder. The digital precoder is designed with joint optimization of intra-group user utility function. It has been shown that more than 50% feedback overhead is reduced by the eigenmodes-based analog precoder design. The spatio-radio resources scheduling and limited feedback-based hybrid precoding increases the sum-rate by 27.6% compared to the sum-rate of one-group case at the cost of 33.3% extra feedback overhead, and reduces the feedback overhead by 62.5% at the cost of 25.2% reduction in sum-rate, compared to the full CSI feedback.

**Author Contributions:** H.K. and I.A., contributed the key idea and defined the main system model. G.B. and M.A. assisted with the system model and the mathematical analysis. I.A. developed the problem formulation.

G.B. and H.K. developed the model for users grouping. H.K., I.A., G.B. and M.A. analysed all the results and added the relevant discussions. All authors contributed to the paper write up.

**Funding:** This research was supported by King Abdul Aziz City for Science and Technology Project under Grant PC-37-66.

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

## Abbreviations

The following abbreviations are used in this manuscript:

ADC	Analog-to-digital converter
AWGN	Additive white Gaussian noise
BS	Base station
CDF	Cumulative distribution function
CSI	Channel state information
DAC	Digital-to-analog converter
DFT	Discrete Fourier transforms
DoA	Direction of arrival
DoD	Direction of departure
DPC	Dirty paper coding
FDD	Frequency division duplex
HBwS	Hybrid beamforming with switches
LOS	Line-of-sight
MIMO	Multiple-input multiple-output
MIP	Mixed integer programming
ML	Machine learning
MMSE	Minimum mean squared error
MPCs	Multi path components
MRT/MRC	Maximum ratio transmission/Maximum ratio combining
MU-MIMO	Multi-User MIMO
OFDM	Orthogonal frequency division multiplexing
RBD	Regularized block diagonalization
RF	Radio frequency
SC	Subspace construction
SINR	Signal-to-interference-plus-noise ratio
SVD	Singular value decomposition
SVM	Support vector machine
TDD	Time division duplex
UE	User equipment
ULA	Uniform linear array

## References

1. 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. [[CrossRef](#)]
2. Kuttu, S.; Member, S.; Sen, D. Beamforming for millimeter wave communications: An inclusive survey. *IEEE Commun. Surv. Tutor.* **2016**, *18*, 949–973. [[CrossRef](#)]
3. Nam, J.; Adhikary, A.; Ahn, J.Y.; Caire, G. Joint Spatial Division and Multiplexing: Opportunistic Beamforming, User Grouping and Simplified Downlink Scheduling. *IEEE J. Sel. Top. Signal Process.* **2014**, *8*, 876–890. [[CrossRef](#)]
4. Adhikary, A.; Al Safadi, E.; Samimi, M.; Wang, R.; Caire, G.; Rappaport, T.; Molisch, A. Joint Spatial Division and Multiplexing for Wave Channels. *IEEE J. Sel. Areas Commun.* **2014**. [[CrossRef](#)]
5. Li, Z.; Han, S.; Sangodoyin, S.; Wang, R.; Molisch, A.F. Joint Optimization of Hybrid Beamforming for Multi-User Massive MIMO Downlink. *IEEE Trans. Wirel. Commun.* **2018**, *1276*, 1–14. [[CrossRef](#)]

6. Li, Z.; Han, S.; Molisch, A.F. Optimizing Channel-Statistics-Based Analog Beamforming for Millimeter-Wave Multi-User Massive MIMO Downlink. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 4288–4303. [[CrossRef](#)]
7. Ratnam, V.; Molisch, A.; Bursalioglu, O.Y.; Papadopoulos, H.C. Hybrid Beamforming with Selection for Multi-user Massive MIMO Systems. *IEEE Trans. Signal Process.* **2018**, *66*, 4105–4121. [[CrossRef](#)]
8. Liu, A.; Lau, V. Phase only RF precoding for massive MIMO systems with limited RF chains. *IEEE Trans. Signal Process.* **2014**, *62*, 4505–4515. [[CrossRef](#)]
9. Li, Z.; Han, S.; Molisch, A.F. User-Centric Virtual Sectorization for Millimeter-Wave Massive MIMO Downlink. *IEEE Trans. Wirel. Commun.* **2018**, *17*, 445–460. [[CrossRef](#)]
10. Ahmed, I.; Khammari, H.; Shahid, A.; Musa, A.; Kim, K.S.; Poorter, E.D.; Moerman, I. A Survey on Hybrid Beamforming Techniques in 5G: Architecture and System Model Perspectives. *IEEE Commun. Surv. Tutor.* **2018**, *20*, 3060–3097. [[CrossRef](#)]
11. Park, S.; Alkhateeb, A.; Heath, R.W. Dynamic Subarrays for Hybrid Precoding in Wideband mmWave MIMO Systems. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 2907–2920. [[CrossRef](#)]
12. Zhu, D.; Li, B.; Liang, P. A Novel Hybrid Beamforming Algorithm with Unified Analog Beamforming by Subspace Construction Based on Partial CSI for Massive MIMO-OFDM Systems. *IEEE Trans. Commun.* **2017**, *65*, 594–607. [[CrossRef](#)]
13. Jeon, Y.; Song, C.; Lee, S.R.; Maeng, S.; Jung, J.; Lee, I. New Beamforming Designs for Joint Spatial Division and Multiplexing in Large-Scale MISO Multi-User Systems. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 3029–3041. [[CrossRef](#)]
14. Jiang, Z.; Chen, S.; Zhou, S. Joint User Scheduling and Beam Selection Optimization for Beam-Based Massive MIMO Downlinks. *IEEE Trans. Wirel. Commun.* **2018**, *17*, 2190–2204. [[CrossRef](#)]
15. Ahmed, I.; Khammari, H.; Shahid, A. Resource allocation for transmit hybrid beamforming in decoupled millimeter wave multiuser-MIMO downlink. *IEEE Access* **2017**, *5*, 170–182. [[CrossRef](#)]
16. O’Shea, T.J.; Hoydis, J. An Introduction to Deep Learning for the Physical Layer. *IEEE Trans. Cogn. Commun. Netw.* **2017**, *3*, 563–575. [[CrossRef](#)]
17. Kim, K.; Lee, J.; Choi, J. Deep Learning Based Pilot Allocation Scheme (DL-PAS) for 5G Massive MIMO System. *IEEE Commun. Lett.* **2018**, *22*, 828–831. [[CrossRef](#)]
18. Wen, C.K.; Shih, W.T.; Jin, S. Deep Learning for Massive MIMO CSI Feedback. *IEEE Wirel. Commun. Lett.* **2018**, *1–4*. [[CrossRef](#)]
19. Joung, J. Machine Learning-Based Antenna Selection in Wireless Communications. *IEEE Commun. Lett.* **2016**, *20*, 2241–2244. [[CrossRef](#)]
20. Wang, J.B.; Wang, J.; Wu, Y.; Wang, J.Y.; Zhu, H.; Lin, M.; Wang, J. A Machine Learning Framework for Resource Allocation Assisted by Cloud Computing. *IEEE Netw.* **2018**, *32*, 144–151. [[CrossRef](#)]
21. Calabrese, F.D.; Wang, L.; Ghadimi, E.; Peters, G.; Soldati, P. Learning Radio Resource Management in 5G Networks: Framework, Opportunities and Challenges. *arXiv* **2016**, arXiv:1611.10253.
22. Ahmed, I.; Khammari, H. Joint Machine Learning based Resource Allocation and Hybrid Beamforming Design for Massive MIMO Systems. In Proceedings of the IEEE Globecom Workshops (GC Wkshps), Abu Dhabi, UAE, 9–13 December 2018; pp. 1–6.
23. Rahman, M.S.; Josiam, K. Low complexity RF beam search algorithms for millimeter-wave systems. In Proceedings of the 2014 IEEE Global Communications Conference (GLOBECOM 2014), Austin, TX, USA, 8–12 December 2014; pp. 3815–3820. [[CrossRef](#)]
24. Almers, P.; Bonek, E.; Burr, A.; Czink, N.; Debbah, M.; Degli-Esposti, V.; Hofstetter, H.; Kyösti, P.; Laurenson, D.; Matz, G.; et al. Survey of channel and radio propagation models for wireless MIMO systems. *Eurasip J. Wirel. Commun. Netw.* **2007**, *2007*, 019070. [[CrossRef](#)]
25. Andrews, J.G.; Bai, T.; Kulkarni, M.; Alkhateeb, A.; Gupta, A.; Heath, R.W. Modeling and Analyzing Millimeter Wave Cellular Systems. *IEEE Trans. Commun.* **2017**, *65*, 403–430. [[CrossRef](#)]
26. Ayach, O.E.; Rajagopal, S.; Abu-Surra, S.; Pi, Z.; Heath, R.W. Spatially sparse precoding in millimeter wave MIMO systems. *IEEE Trans. Wirel. Commun.* **2014**, *13*, 1499–1513. [[CrossRef](#)]
27. Sohrabi, F.; Yu, W. Hybrid digital and analog beamforming design for large-scale MIMO systems. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, Australia, 19–24 April 2015; pp. 2929–2933. [[CrossRef](#)]

28. Ayach, O.E.; Heath, R.W.; Abu-Surra, S.; Rajagopal, S.; Pi, Z. Low complexity precoding for large millimeter wave MIMO systems. In Proceedings of the IEEE International Conference on Communications (ICC), Ottawa, ON, Canada, 10–15 June 2012; pp. 3724–3729. [[CrossRef](#)]
29. Rappaport, T.S.; Gutierrez, F.; Ben-Dor, E.; Murdock, J.N.; Qiao, Y.; Tamir, J.I. Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor Urban cellular communications. *IEEE Trans. Antennas Propag.* **2013**, *61*, 1850–1859. [[CrossRef](#)]
30. Nam, J.; Ahn, J.Y.; Adhikary, A.; Caire, G. Joint spatial division and multiplexing: Realizing massive MIMO gains with limited channel state information. In Proceedings of the 2012 46th Annual Conference on Information Sciences and Systems (CISS 2012), Princeton, NJ, USA, 21–23 March 2012; Volume 1. [[CrossRef](#)]
31. Lu, C.; Wang, W.; Zhong, W.; Gao, X. User scheduling and beam allocation for massive MIMO systems with two-stage precoding. In Proceedings of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Valencia, Spain, 4–8 September 2016. [[CrossRef](#)]
32. Andrews, M. A survey of scheduling theory in wireless data networks. In *Wireless Communications*; Springer: New York, NY, USA, 2007; pp. 1–17.
33. Escudero-Garzas, J.J.; Devillers, B.; Garcia-Armada, A. Fairness-Adaptive Goodput-Based Resource Allocation in OFDMA Downlink with ARQ. *IEEE Trans. Veh. Technol.* **2014**, *63*, 1178–1192. [[CrossRef](#)]
34. Alkhateeb, A., Jr.; Heath, R.W. Frequency Selective Hybrid Precoding for Limited Feedback Millimeter Wave Systems. *IEEE Trans. Commun.* **2016**, *64*, 1801–1818. [[CrossRef](#)]
35. Kang, J.; Lee, J.H.; Choi, W. Dimension-optimized Two-Stage Precoder Design for Massive MIMO Systems with Limited Feedback. In Proceedings of the IEEE Military Communications Conference MILCOM, Los Angeles, CA, USA, 29–31 October 2018; pp. 530–535. [[CrossRef](#)]
36. Tataria, H.; Smith, P.J.; Greenstein, L.J.; Dmochowski, P.A.; Shafi, M. Performance and analysis of downlink multiuser MIMO systems with regularized zero-forcing precoding in Ricean fading channels. In Proceedings of the IEEE International Conference on Communications, Kuala Lumpur, Malaysia, 22–27 May 2016; pp. 1–7.
37. Ho, W.; Liang, Y. Optimal Resource Allocation for Multiuser MIMO-OFDM Systems With User Rate Constraints. *IEEE Trans. Veh. Technol.* **2009**, *58*, 1190–1203. [[CrossRef](#)]
38. Spencer, Q.; Swindlehurst, A.; Haardt, M. Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels. *IEEE Trans. Signal Process.* **2004**, *52*, 461–471. [[CrossRef](#)]
39. Li, Z.; Han, S.; Molisch, A.F. Hybrid beamforming design for millimeter-wave multi-user massive MIMO downlink. In Proceedings of the 2016 IEEE International Conference on Communications (ICC 2016), Kuala Lumpur, Malaysia, 22–27 May 2016. [[CrossRef](#)]
40. Kang, J.; Lee, J.H.; Choi, W. Machine Learning-Based Dimension Optimization for Two-Stage Precoder in Massive MIMO Systems with Limited Feedback. *Appl. Sci.* **2019**, *9*, 2894. [[CrossRef](#)]

