MU-MIMO with Type-II Codebook Precoding and User Grouping in ns-3 5G-LENA

Amirhassan Babazadeh Darabi
RWTH Aachen University
Aachen, Germany
amirhassan.babazadeh@mcc.rwth-aachen.de

Marina Petrova
RWTH Aachen University
Aachen, Germany
petrova@mcc.rwth-aachen.de

Ljiljana Simić
RWTH Aachen University
Aachen, Germany
simic@mcc.rwth-aachen.de

***Abstract*—Multi-user multiple-input multiple-output (MU-MIMO) is a key technique for improving spectral efficiency in 5G and beyond networks by allowing multiple users to be served simultaneously over the same time-frequency resources. Realizing downlink MU-MIMO, however, requires joint precoder selection and user grouping that accounts for the spatial compatibility among co-scheduled users. Existing grouping schemes largely rely on explicit (full) channel state information (CSI) and are thus not compatible with 3GPP codebook-based MU-MIMO that uses implicit CSI feedback to select predefined precoders from Type-I or Type-II codebooks. In this paper, we propose *CPOG*, a novel user grouping scheme based on codebook precoder orthogonality, which exploits the Type-I/II codebooks to capture each user's dominant spatial subspace and group spatially compatible users. We also present and open-source an extension of the ns-3 5G-LENA module to enable downlink MU-MIMO transmission, incorporate Type-II codebooks, and implement MU-MIMO user grouping and scheduling. We present extensive full-stack ns-3 simulation results comparing codebook-based MU-MIMO with baseline full-CSI MU-MIMO schemes from the literature, demonstrating the tradeoff between throughput performance and signaling overhead/complexity. Importantly, we show that the proposed low-complexity CPOG user grouping outperforms the k-means clustering benchmark in terms of MU-MIMO throughput and spatial multiplexing efficiency.**

***Index Terms*—downlink MU-MIMO, codebook-based precoding, CSI feedback, system-level simulation, ns-3**

I. Introduction

The use of multiple antennas at both the transmitter and receiver, commonly referred to as multiple-input multiple-output (MIMO), has been a key enabler of high spectral efficiency in LTE and 5G New Radio (NR), and is expected to play an even greater role in next-generation networks, where the spatial dimension is scaled up through massive MIMO (mMIMO) and cell-free massive MIMO (CF-mMIMO) deployments [1], [2]. In multi-user MIMO (MU-MIMO), the gNB exploits spatial multiplexing to transmit multiple data streams and serve multiple users at once on the same time-frequency resources [3], thus increasing system capacity.

Realizing these MU-MIMO gains, however, depends on the accuracy of the channel state information (CSI) available at the transmitter. Acquiring the full channel matrix is costly, since the feedback signaling overhead grows with the number of antennas and antenna ports. This makes explicit full-CSI acquisition prohibitive in practice, even more so in the mMIMO regime where MU-MIMO is most beneficial. To reduce this cost, 5G-NR relies on standardized implicit feedback defined by 3GPP [4], where the UE reports the precoding matrix indicator (PMI) and rank indicator (RI) from which the gNB reconstructs a precoding matrix. To this end, two codebook types are specified, Type-I and Type-II, differing in channel feedback granularity. Explicit CSI feedback would allow the gNB to form an optimal precoder with the full channel matrix to make the users' channels orthogonal to each other and suppress the MU-MIMO inter-user interference. By contrast, codebook-based MU-MIMO can only select the precoder from a predefined set to orthogonalize the users. The Type-II codebook is considered a better solution for downlink MU-MIMO than Type-I [5], as it provides finer resolution and richer spatial information that MU-MIMO requires to suppress inter-user interference, albeit at the cost of higher feedback signaling overhead and greater precoder-selection signal processing complexity at the user side.

Although MU-MIMO precoding, such as zero forcing (ZF), is designed to suppress inter-user interference, it alone cannot guarantee efficient spatial multiplexing, as its performance strongly depends on the spatial correlation of the users' channels. User grouping is therefore introduced to reduce the residual inter-user interference after precoding [5]. Explicit CSI methods use various techniques to group the users; singular value decomposition (SVD) [6], the orthogonal projection matrix [7], or the QR decomposition [8] have been introduced in the literature to group the spatially compatible users and increase the MU-MIMO gains. In addition, spatial clustering approaches have been proposed to group the users according to the correlation of their effective channels [9], commonly through unsupervised methods such as k-means clustering of the dominant channel directions [10]. However, these methods heavily depend on the full CSI feedback and are not directly compatible with codebook-based precoders to allocate user groups. Namely, user grouping is key to realizing codebook-based MU-MIMO, but is largely unaddressed in the literature. To address this gap, in this paper, we propose *CPOG* (codebook precoder orthogonality grouping), a novel user grouping method for codebook-based MU-MIMO precoding. CPOG groups the spatially compatible users based on their precoder orthogonality as a proxy for the orthogonality of their channels, and allocates shared resources to group members on a separate spatial layer.

979-8-3195-0167-7/26/$31.00 ©2026 IEEE

To assess the practical performance and feasibility of candidate MU-MIMO schemes in 5G-and-beyond networks, system-level evaluation is of utmost importance, since spatial multiplexing gains do not arise solely from PHY operations; rather, they also depend on how the scheduler exploits spatial compatibility by user grouping and how this interacts with link adaptation and higher-layer traffic. Network simulators such as ns-3 [11] enable such full-stack evaluation by jointly modeling the PHY, MAC, and upper layers under realistic 3GPP 5G-NR scenarios, capturing the coupling between precoder selection and multi-user grouping that a PHY-only abstraction would miss. However, although the existing ns-3 5G-LENA module [12] supports SU-MIMO and implements Type-I codebook precoding, it supports neither MU-MIMO nor Type-II codebooks. The authors in [13] extend 5G-LENA to implement various precoder selection approaches to reduce computational complexity at a marginal throughput loss, but also support only Type-I codebooks. The authors in [14] implement MU-MIMO via hybrid beamforming in the ns-3 mmWave module, a legacy predecessor of 5G-LENA; however, the implementation is specific to the mm-wave channel, implements PHY-only multi-user interference suppression without considering user grouping, and does not consider codebook-based precoding. To the best of our knowledge, neither Type-II codebook-based precoding with multi-user grouping has been implemented in 5G-LENA, nor has a full-stack comparative evaluation of codebook-based multi-user grouping methods with full CSI been conducted in this context.

In this paper, we address this gap by presenting an open-source extension of 5-LENA to support downlink MU-MIMO transmission, Type-II codebooks, and user grouping for spatially compatible users. We use our ns-3 framework to present, to the best of our knowledge, the first extensive system-level simulation results comparing codebook-based and explicit CSI MU-MIMO in 5G-NR, thereby quantifying the trade-off between throughput performance and signaling overhead and computational complexity. Importantly, we show that for codebook-based MIMO, the proposed CPOG user grouping outperforms the k-means clustering benchmark in terms of both higher MU-MIMO throughput and spatial multiplexing and lower computational complexity, making it a practical 3GPP-compliant scheme for 5G-and-beyond MU-MIMO.

II. System Model

We consider a 5G-NR downlink MU-MIMO system with a single gNB, equipped with a dual-polarized uniform planar array (UPA) with N_t antenna ports. The user equipments (UEs) have N_r antenna ports and measure the channel via receiving channel state information reference signals (CSI-RSs) sent from the gNB. The channel bandwidth is split into resource blocks (RBs), following 3GPP NR. The received signal per resource element (RE), i.e., per subcarrier and per OFDM symbol, $\mathbf{y} \in \mathbb{C}^{N_r \times 1}$ at the m-th UE, with total M co-scheduled users, after the spatial whitening of the noise-plus-interference, is written as

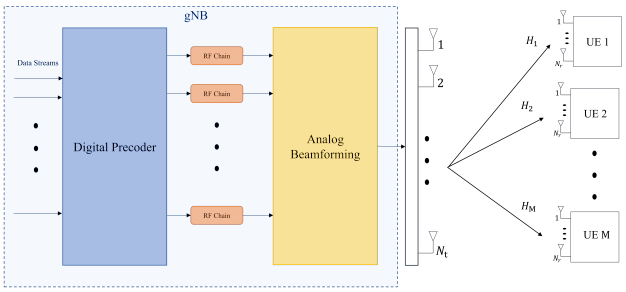


Fig. 1. Downlink MU-MIMO system model.

$$\mathbf{y}_m = \mathbf{H}_m \mathbf{W}_m \mathbf{x}_m + \mathbf{H}_m \sum_{\substack{1 \le j \le M \\ j \ne m}} \mathbf{W}_j \mathbf{x}_j + \mathbf{n}_m, \tag{1}$$

where $\mathbf{H}_m \in \mathbb{C}^{N_r \times 2N_t}$ is the downlink channel matrix of the UE m, $\mathbf{W}_m \in \mathbb{C}^{2N_t \times \nu_m}$ is the precoding matrix of UE m, and $\mathbf{x}_m \in \mathbb{C}^{\nu_m \times 1}$ is the transmit signal vector with $\nu_m \ge 1$ being the number of spatial layers. Fig. 1 illustrates the system model. We assume a linear receiving filter for UE m, $\mathbf{F}_m$, and the estimated symbol at the UE side is obtained as

$$\begin{aligned} \hat{\mathbf{x}}_m &= \mathbf{F}_m^H \mathbf{y}_m \\ &= \mathbf{F}_m^H \mathbf{H}_m \mathbf{W}_m \mathbf{x}_m + \mathbf{F}_m^H \mathbf{H}_m \sum_{\substack{1 \le j \le M \\ j \ne m}} \mathbf{W}_j \mathbf{x}_j + \mathbf{F}_m^H \mathbf{n}_m, \end{aligned} \tag{2}$$

where the first term is the desired signal, the second term is the residual inter-user interference, and the third term is the residual noise. $\mathbf{F}_m^H$ is the linear minimum mean square (LMMSE) equalized filter and is written as

$$\begin{aligned} \mathbf{F}_m^H &= \mathbf{W}_m^H \mathbf{H}_m^H \Big[\mathbf{H}_m \mathbf{W} \mathbf{W}^H \mathbf{H}_m^H + N_0 \mathbf{I} \Big]^{-1} \\ &= \mathbf{W}_m^H \mathbf{H}_m^H \mathbf{R}_m^{-1}, \end{aligned} \tag{3}$$

where $\mathbf{R}_m = \mathbf{H}_m \Big(\sum_{j=1}^{M} \mathbf{W}_j \mathbf{W}_j^H \Big) \mathbf{H}_m^H + N_0 \mathbf{I}$, and $\mathbf{W}$ is the stacked precoders of the M UEs, $\mathbf{I}$ is the identity matrix, and N_0 is the noise power. The estimated signal for UE m is written as

$$\begin{aligned} \hat{\mathbf{x}}_m &= \mathbf{W}_m^H \mathbf{H}_m^H \mathbf{R}_m^{-1} \mathbf{H}_m \mathbf{W}_m \mathbf{x}_m \\ &\quad + \mathbf{W}_m^H \mathbf{H}_m^H \mathbf{R}_m^{-1} \mathbf{H}_m \sum_{\substack{1 \le j \le M \\ j \ne m}} \mathbf{W}_j \mathbf{x}_j + \mathbf{W}_m^H \mathbf{H}_m^H \mathbf{R}_m^{-1} \mathbf{n}_m. \end{aligned} \tag{4}$$

The signal-to-interference-plus-noise ratio (SINR) measured at the receiver for the UE m is

$$\gamma_m(\mathbf{H}_m, \mathbf{W}_m) = \frac{(\mathbf{H}_m \mathbf{W}_m)^H \mathbf{R}_m^{-1} (\mathbf{H}_m \mathbf{W}_m)}{1 - (\mathbf{H}_m \mathbf{W}_m)^H \mathbf{R}_m^{-1} (\mathbf{H}_m \mathbf{W}_m)}, \quad \mathbf{W}_m \in \mathbf{W}, \tag{5}$$

where $\mathbf{W}_m$ is the m-th column of the precoding matrix $\mathbf{W}$.

In the 3GPP standard, the concept of a subband (SB) has been introduced to partition the allocated bandwidth into multiple subbands, each consisting of up to N_{SB} physical resource blocks (PRBs). Since channel variations across frequency are typically limited, the precoder can be assumed constant over all PRBs within a given SB for a specific OFDM symbol. Consequently, a representative channel is considered for each SB t to compute the corresponding precoder, denoted as $\mathbf{H}(t)$ [15]. The downsampling method is the averaged channel over the SB.

Type-I and Type-II codebooks are standardized in 3GPP to facilitate implicit channel feedback [4]. In the lower-resolution Type-I codebook, each spatial layer is formed from a single IDFT beam together with a co-phasing coefficient between the two polarizations. The rank-1 precoder at SB t is written as

$$\mathbf{w}_t = \frac{1}{\sqrt{2N_t}} \begin{bmatrix} \mathbf{v}_{m_1 m_2} \\ \varphi_t \, \mathbf{v}_{m_1 m_2} \end{bmatrix}, \tag{6}$$

where $\mathbf{v}_{m_1 m_2}$ is the selected IDFT beam and φ_t is the inter-polarization co-phasing coefficient. Since only one beam is reported per layer, Type-I incurs low feedback overhead but yields a coarse, single-beam approximation of the channel direction. It also supports higher transmission ranks ν (up to rank 8 for a single panel). The higher-resolution Type-II codebook is restricted to $\nu \leq 2$, but reconstructs the dominant channel subspace far more accurately by linearly combining $L \geq 2$ beams [15]. The Type-II codebook precoding matrix for ranks 1 and 2 at the t-th SB, as Table 5.2.2.2.3-5 in [4], is written as

$$\mathbf{W}_t^1 = \mathbf{w}_t^{(1)}, \quad \mathbf{W}_t^2 = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{w}_t^{(1)} & \mathbf{w}_t^{(2)} \end{bmatrix}. \tag{7}$$

Consequently, the Type-II precoder for a given spatial layer and SB can be represented as the product $\mathbf{W}_t^{(l)} = \mathbf{W}_1 \mathbf{w}_2$ where

$$\mathbf{W}_1 = \begin{bmatrix} \mathbf{v}_{m_1^{(0)} m_2^{(0)}} & \cdots & \mathbf{v}_{m_1^{(L-1)} m_2^{(L-1)}} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{v}_{m_1^{(0)} m_2^{(0)}} & \cdots & \mathbf{v}_{m_1^{(L-1)} m_2^{(L-1)}} \end{bmatrix}. \tag{8}$$

The matrix $\mathbf{W}_1 \in \mathbb{C}^{2N_t \times 2L}$ determines the L chosen IDFT beams $\mathbf{v}_{m_1^{(i)} m_2^{(i)}}$ for both polarizations. The L beams are chosen as

$$\mathbf{v}_{l^*,m^*} = \underset{\mathbf{v}_{l,m} \in \mathcal{B}}{\operatorname{argmax}} \, \mathbf{v}_{l,m}^H (\mathbf{R}_{11} + \mathbf{R}_{22}) \mathbf{v}_{l,m}, \tag{9}$$

where $\mathcal{B}$ is the IDFT grid of the beams and $\mathbf{R}_{11}$ and $\mathbf{R}_{22}$ are obtained from

$$\begin{aligned} \mathbf{R} &= \frac{1}{N_3} \sum_{t=0}^{N_3-1} \mathbf{H}^H(t) \mathbf{H}(t) \\ &= \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{bmatrix}. \end{aligned} \tag{10}$$

$\mathbf{w}_2$ is a vector of beam combining coefficients (BCCs) defined for each SB and spatial layer. More precisely, for a given SB t and layer l, the precoder can be written as

$$\mathbf{w}_t^{(l)} = \Lambda_t^{(l)} \mathbf{W}_1 \mathbf{W}_{C1}^{(l)} \mathbf{w}_{C2,t}^{(l)}, \tag{11}$$

where $\mathbf{W}_{C1}^{(l)} \in \mathbb{C}^{2L \times 2L}$ is a diagonal matrix containing the quantized wideband (WB) amplitudes, while $\mathbf{w}_{C2,t}^{(l)} \in \mathbb{C}^{2L}$ contains the quantized amplitudes and phases for SB t. Finally, $\Lambda_t^{(l)}$ is the normalization factor. The values for wideband amplitudes, subband amplitudes, and subband phases are selected from Table 5.2.2.2.3-2 and 5.2.2.2.3-3 of [4].

The Type-II precoder can also be represented across multiple SBs. For N_3 SBs and layer l, the concatenated precoder is given by

$$\mathbf{W}^{(l)} = \left[\mathbf{w}_0^{(l)}, \ldots, \mathbf{w}_{N_3-1}^{(l)} \right]. \tag{12}$$

Using the Type-II decomposition, this can be written as

$$\mathbf{W}^{(l)} = \mathbf{W}_1 \mathbf{W}_{C1}^{(l)} \left[\mathbf{w}_{C2,0}^{(l)}, \ldots, \mathbf{w}_{C2,N_3-1}^{(l)} \right] \mathbf{\Lambda}^{(l)}, \tag{13}$$

or equivalently,

$$\mathbf{W}^{(l)} = \mathbf{W}_1 \mathbf{W}_{C1}^{(l)} \mathbf{W}_{C2}^{(l)} \mathbf{\Lambda}^{(l)}. \tag{14}$$

Here, $\mathbf{W}_{C2}^{(l)}$ contains the quantized subband coefficients of all N_3 SBs, and

$$\mathbf{\Lambda}^{(l)} = \operatorname{diag}\left(\left[\Lambda_0^{(l)}, \ldots, \Lambda_{N_3-1}^{(l)} \right] \right) \tag{15}$$

collects the normalization factors of all subband precoders. The t-th column of $\mathbf{W}^{(l)}$ therefore corresponds to the precoder applied to the t-th SB of layer l.

Designing the optimal MU-MIMO precoder matrix tries to achieve spatial multiplexing, increase link reliability, and mitigate the residual inter-user interference term in (2) [5]. The stacked received signal of M UEs, $\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \, \mathbf{y}_2 \cdots \mathbf{y}_M \end{bmatrix}^T \in \mathbb{C}^{MN_r \times 1}$, is written as

$$\mathbf{y} = \begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_M \end{bmatrix} \begin{bmatrix} \mathbf{W}_1 \, \ldots \, \mathbf{W}_M \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_M \end{bmatrix} + \begin{bmatrix} \mathbf{n}_1 \\ \vdots \\ \mathbf{n}_M \end{bmatrix}. \tag{16}$$

SVD is utilized to factorize the channel matrix as $\mathbf{H}_m = \mathbf{U}_m \mathbf{\Sigma}_m \mathbf{V}_m^H$, where $\mathbf{U}_m \in \mathbb{C}^{N_r \times N_r}$ and $\mathbf{V}_m \in \mathbb{C}^{2N_t \times N_r}$ are unitary matrices, and the diagonal matrix $\mathbf{\Sigma}_m \in \mathbb{C}^{N_r \times N_r}$ contains the sorted real positive eigenvalues. By applying ZF to the matrix $\mathbf{Z} = \begin{bmatrix} \mathbf{V}_1^H \\ \vdots \\ \mathbf{V}_M^H \end{bmatrix}$, we can determine the final precoding matrix. Let $\mathbf{P}^{\text{ZF}}$ be the pseudo inverse of the matrix $\mathbf{Z}$ such that $\mathbf{P}^{\text{ZF}} = \mathbf{Z}^H \left(\mathbf{Z}\mathbf{Z}^H \right)^{-1} = \left[\mathbf{P}_1^{\text{ZF}} \ldots \mathbf{P}_M^{\text{ZF}} \right]$. The precoding matrix corresponding to the m-th UE can be inferred from the first ν_m columns of the $\mathbf{P}_m^{\text{ZF}}$ matrix, where ν_m is the transmission rank of the UE m.

Since the columns of the Type-II precoder reported by a given UE k can be viewed as a close approximation of the first ν_k columns of $\mathbf{V}_k$ (associated with the dominant

eigenvalues of the channel) [15], the MU-MIMO precoder can be constructed by stacking the individual precoders of the selected UEs. In contrast, the Type-I precoder is generally considered insufficiently accurate for this approximation.

III. Proposed CPOG User Grouping Scheme for Codebook-Based MU-MIMO

A. The CPOG Algorithm

In the proposed CPOG (codebook precoder orthogonality grouping) scheme, user grouping is performed at each scheduling instant following a greedy, priority-ordered procedure. Candidate users are ordered according to the scheduling metric (e.g., round-robin or proportional-fair), so that higher-priority users are served first. A new group is initialized with the highest-priority user that still has data to transmit, and the gNB has sufficient resources. Each subsequent candidate is then examined in order and admitted to the group only if (i) it has sufficient resources and (ii) its precoding vector is deemed to be compatible with the precoding vectors of *all* users already in the group. Performing the check against every existing member, rather than only against the first user, ensures that compatibility holds across the entire group. Admission continues until no further compatible users can be added or the group reaches the maximum number of spatial layers L_{max} available for transmission. The procedure then repeats over the remaining users to form the next group.

The compatibility between two users is determined in CPOG by the degree of orthogonality between their precoding vectors. Let $\mathbf{W}_i \in \mathbb{C}^{2N_t \times 1}$ and $\mathbf{W}_j \in \mathbb{C}^{2N_t \times 1}$ denote the unit-norm rank-1 codebook-based precoding vectors of user i and user j, respectively. Two users are ideally compatible when their precoders are mutually orthogonal,

$$\mathbf{W}_i^H \mathbf{W}_j = 0, \quad i \neq j, \tag{17}$$

in which case the stream of user i causes no interference to user j and the subsequent ZF precoder incurs no power penalty. In practice, exact orthogonality is rarely achievable due to channel estimation errors, the finite resolution of the codebook, and the limited spatial degrees of freedom of the array. The residual correlation is captured by the inner product $\mathbf{W}_i^H \mathbf{W}_j$, whose magnitude reflects the inter-user interference that remains after precoding. We therefore adopt the scalar compatibility metric

$$\eta_{ij} = |\mathbf{W}_i^H \mathbf{W}_j|, \tag{18}$$

which, for unit-norm precoders, satisfies $\eta_{ij} \in [0, 1]$ by the Cauchy–Schwarz inequality: $\eta_{ij} = 0$ denotes orthogonal (fully compatible) precoders and $\eta_{ij} = 1$ denotes fully aligned ones. Two users are declared compatible, and hence eligible to share the same resources, when

$$\eta_{ij} = |\mathbf{W}_i^H \mathbf{W}_j| \leq \delta_{\text{thr}}, \quad i \neq j, \tag{19}$$

where δ_{thr} is the compatibility threshold.

The threshold δ_{thr} governs the trade-off between spatial multiplexing gain and residual inter-user interference. A small δ_{thr} enforces near-orthogonality, keeping the post-precoding inter-user interference and the power penalty of zero-forcing low, but admits fewer users per group and thus limits the multiplexing gain. Conversely, a larger δ_{thr} allows more UEs to be co-scheduled at the expense of higher residual interference. The value of δ_{thr}, together with the limit on the group size L_{max} (i.e., the number of co-scheduled users), and is therefore a configurable design parameter in our evaluation that directly shapes the achievable throughput.

Algorithm 1 Proposed CPOG MU-MIMO User Grouping

Require: Active UE set $\mathcal{U}$, compatibility threshold δ_{thr}, maximum layers $L_{\max}$
Ensure: User groups $\mathcal{G}$

Initialize the compatibility threshold δ_{thr}
// UE side: Implicit CSI feedback
for each UE $i \in \mathcal{U}$ **do**
  Estimate the downlink channel $\mathbf{H}_i$
  Compute the precoder $\mathbf{W}_i$ from the codebook
  Report $\mathbf{W}_i$ (PMI, RI, CQI) to the gNB
end for
// gNB side: user grouping
$\mathcal{G} \leftarrow \emptyset$
Order $\mathcal{U}$ by the scheduling metric
for each UE $i \in \mathcal{U}$ **do**
  $isCompatible \leftarrow$ FALSE
  for each group $\mathcal{G}_g \in \mathcal{G}$ **do**
    if $|\mathcal{G}_g| < L_{\max}$ **and** $\eta_{ij} \leq \delta_{\text{thr}}\ \forall j \in \mathcal{G}_g$ **and** resources available **then**
      $\mathcal{G}_g \leftarrow \mathcal{G}_g \cup \{i\}$ ▷ i compatible with all members
      $isCompatible \leftarrow$ TRUE
      break
    end if
  end for
  if $isCompatible =$ FALSE **then**
    $\mathcal{G} \leftarrow \mathcal{G} \cup \{\{i\}\}$ ▷ start a new group
  end if
end for

The complete CPOG MU-MIMO procedure, spanning the implicit CSI feedback at the UE and the user grouping and group-level scheduling at the gNB, is summarized in Algorithm 1. After the compatibility threshold δ_{thr} is initialized, each UE estimates its downlink channel, computes its codebook-based precoder, and reports it to the gNB. The gNB then iterates over the active UEs in priority order and places each UE into a group whose members are all mutually compatible according to the criterion in (19), and creates a new group whenever the UE is incompatible with the existing ones.

B. MU-MIMO Benchmarks & Computational Complexity

To evaluate the proposed CPOG user grouping for codebook-based MU-MIMO, we benchmark it against three representative baseline MU-MIMO schemes that differ in the

TABLE I
COMPUTATIONAL COMPLEXITY OF THE PROPOSED CPOG AND THE BASELINE USER-GROUPING SCHEMES (GNB-SIDE, CSI-FEEDBACK COST EXCLUDED).

Method	Preprocessing	Grouping / clustering	Overall complexity
Proposed CPOG (Type-I & II)	–	$\mathcal{O}(K^2 N_t)$	$\mathcal{O}(K^2 N_t)$
NS-ZF, explicit CSI	$\mathcal{O}(KN_r^2 N_t)$	$\mathcal{O}(K^2 N_t)$	$\mathcal{O}(KN_r^2 N_t + K^2 N_t)$
D-ZF, explicit CSI	$\mathcal{O}(KN_t^2 N_r)$	$\mathcal{O}(K^2 N_t^3)$	$\mathcal{O}(KN_t^2 N_r + K^2 N_t^3)$
k-means, Type-I	–	$\mathcal{O}\big(T(KCN_t + KN_t^2 + CN_t^3)\big)$	$\mathcal{O}\big(T(KCN_t + KN_t^2 + CN_t^3)\big)$
k-means, Type-II	–	$\mathcal{O}\big(T(KCN_t + KN_t^2 + CN_t^3)\big)$	$\mathcal{O}\big(T(KCN_t + KN_t^2 + CN_t^3)\big)$
k-means, explicit CSI	$\mathcal{O}(KN_r^2 N_t)$	$\mathcal{O}\big(T(KCN_t N_r^2 + KN_t^2 N_r + CN_t^3)\big)$	$\mathcal{O}\big(KN_r^2 N_t + T(KCN_t N_r^2 + KN_t^2 N_r + CN_t^3)\big)$

K: number of UEs; N_t/N_r: gNB/UE antennas; C: clusters; T: k-means iterations;

CSI granularity and the metric used to determine user spatial compatibility. The first is *null space projection-based grouping with explicit CSI (**NS-ZF, explicit CSI**)*, which assumes explicit knowledge of the complete downlink channel at the gNB and groups users according to the semi-orthogonality of their full channel vectors [16]. As it assumes full and unquantized CSI, it serves as a performance reference, although the associated feedback overhead makes it impractical in 3GPP networks. The second is *determinant-based grouping with explicit CSI (**D-ZF, explicit CSI**)*, which favors user sets that maximize the determinant, i.e., the Gram volume, of the combined effective channel matrix; a larger determinant indicates more linearly independent and better-conditioned channels, and hence a higher spatial multiplexing capability [17]. The third is ***k-means clustering***, which clusters users in an unsupervised manner based on a spatial-similarity feature, such as channel correlation or precoder direction, and co-schedules users that fall within different clusters [10]. These baselines span the range from full-CSI, capacity-oriented criteria to clustering heuristics, providing a comprehensive benchmark for evaluating the accuracy–overhead trade-off of the proposed codebook-based grouping.

We also derive the complexity of the proposed method and compare it to the baseline schemes in terms of $\mathcal{O}(\cdot)$ as summarized in Table I. Throughout this analysis, we do not account for the CSI feedback transmitted by each UE to the gNB, and we report only the operations performed at the gNB. The proposed CPOG has identical complexity for the Type-I and Type-II codebooks, as it evaluates precoder compatibility via a greedy search regardless of codebook type. The pairwise compatibility test reduces to an inner product of complexity $\mathcal{O}(N_t)$. The greedy search evaluates $\binom{K}{2}$ candidate pairs, i.e. $\mathcal{O}(K^2)$ tests, which yields an overall complexity of $\mathcal{O}(K^2 N_t)$. Crucially, CPOG requires neither a per-UE SVD nor a covariance or determinant evaluation, so its cost is governed solely by this inner-product search. The NS-ZF baseline, instead, operates on the complete channel matrix of each UE. It performs a per-UE SVD at a cost of $\mathcal{O}(KN_r^2 N_t)$ to obtain the dominant singular directions, after which the grouping metric reduces to the same $\mathcal{O}(K^2 N_t)$ inner-product search, giving an overall complexity of $\mathcal{O}(KN_r^2 N_t + K^2 N_t)$. The D-ZF scheme forms the per-UE covariance matrices at a cost of $\mathcal{O}(KN_t^2 N_r)$ and incrementally evaluates the determinant ratio of the aggregate covariance, whose naive evaluation dominates the cost at $\mathcal{O}(K^2 N_t^3)$. Finally, the k-means baselines share a common clustering core of $\mathcal{O}\big(T(KCN_t + KN_t^2 + CN_t^3)\big)$ where C and T are the number of clusters and k-means iterations, respectively. The explicit CSI method incurs an additional per-UE SVD term of $\mathcal{O}(KN_r^2 N_t)$.

IV. IMPLEMENTATION OF MU-MIMO IN NS-3 5G-LENA

This section gives an overview of our implementation of MU-MIMO in ns-3 5G-LENA. Fig. 2 gives an overview of the newly introduced and modified classes marked in **bold** and *italics*, respectively.

The `NrCbTypeTwo` class implements the Type-II codebook analogous to the existing Type-I codebook class. We construct the Type-II CSI feedback, sent to the gNB, after selecting the optimal precoding matrix based on the estimated channel matrix. In the simulator, the CSI feedback generated via the function `NrUePhy::GenerateDlCqiReportMimo()` is sent to the precoder selection method to compute the optimal precoding. The class `NrPmSearchTypeTwo` selects the optimal transmission rank, optimal precoding matrix, the best channel quality information (CQI), and the best modulation and coding scheme (MCS) based on the received feedback through the function `CreateCqiFeedbackMimo()`. In the implementation, first the rank of the precoding matrix is selected in the function `NrPmSearchTypeTwo::SelectRank()`, which cannot exceed rank 2 due to the 3GPP limitation. Then, we choose the optimal precoder for that user and select the rank using the function `NrPmSearchTypeTwo:UpdateAllPrecoding()`.

Conventional schedulers in the ns-3 simulator group users according to their serving beam and prevent UEs associated with different beams from being scheduled on the same time–frequency resources. While this is appropriate for single-user, beam-based transmission, it precludes MU-MIMO, in which spatially separable users must share the same resources on distinct spatial layers. We remove this restriction so that all active UEs are treated as candidates for co-scheduling regardless of their serving beam, and compatibility is assessed directly from their reported precoding vectors rather than from beam alignment. Since each precoding vector already captures the dominant channel direction reported by the UE, two users

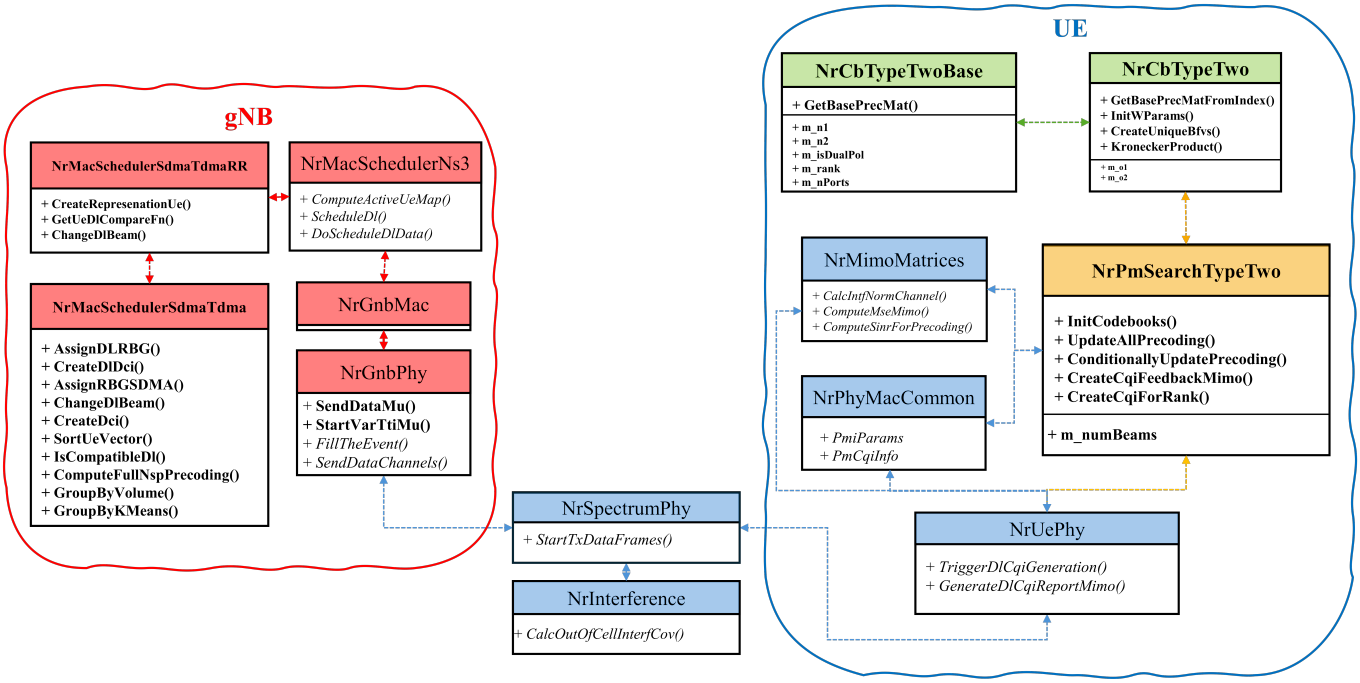


Fig. 2. Diagram of our downlink MU-MIMO implementation in ns-3 5G-LENA, with new functions marked in **bold**, and modified functions in *italics*.

whose precoders are nearly orthogonal can be separated by linear precoding with little residual interference. We implement the `NrMacSchedulerSdmaTdma` class that is responsible for grouping spatially compatible users and implements Algorithm 1. Specifically, the function `IsCompatibleDl()` implements the CPOG MU-MIMO compatibility criteria.

Once the compatible groups are formed, the scheduler allocates resources at the group level rather than on a per-UE basis. Each group is treated as a single scheduling entity to which time–frequency resources, i.e., symbols and resource block groups (RBGs), are assigned. Within a group, the co-scheduled UEs share the same resources on distinct spatial layers, and the final transmit precoder for that allocation is obtained by stacking the per-UE precoders and applying zero-forcing to null the inter-user interference across the group. Finally, the scheduler creates the downlink control information (DCI) corresponding to each UE in the group. Algorithm 2 gives the per-group scheduling method extended within `NrMacSchedulerNs3` class.

Our MU-MIMO scheduler has a round-robin discipline and adopts a hybrid strategy as follows. The available symbols within a slot are first partitioned across the different user groups in a TDMA fashion; within each allocated symbol, the compatible UEs of a group are then multiplexed in the spatial domain. This design preserves the simplicity and fairness of TDMA scheduling while exploiting the spatial degrees of freedom offered by MU-MIMO. When no spatially separable users are available to form a group, the scheme naturally falls back to conventional TDMA scheduling.

In addition, we modify the `NrGnbPhy` class to enable simultaneous multi-user transmission. Specifically, once the MAC layer forwards the slot scheduling decision to the gNB PHY layer, the class constructs the user groups based on their DCI and forwards the corresponding multi-user data channels to `NrSpectrumPhy` to initiate multi-user data transmission. We plan to release our implementation of the MU-MIMO framework via [18] as an open-source code once a stable version is approved.

The baseline methods (*cf.* Section III-B) are implemented in the `NrMacSchedulerSdmaTdma` class, which utilizes the full channel information to group active users in that transmission instance. We implement the functions `ComputeFullNspPrecoding`, `GroupByVolume`, and `GroupByKMeans` to compute the precoding matrices and group the users based on NS-ZF, D-ZF, and k-means clustering, respectively. These functions use the full channel matrix of each user at the gNB side to perform the MU-MIMO user grouping algorithm. The `GroupByKMeans` uses the dominant channel direction and creates L_{max} distinct groups based on their chordal distance. Each group contains the most similar users that cannot be scheduled at the same time, and the users in different groups are scheduled together.

V. Results

We evaluate the performance of codebook-based MU-MIMO with our proposed CPOG user grouping (*cf.* Section III) against benchmark MU-MUMO schemes in Section V-B, via ns-3 simulations as outlined in Section V-A.

Algorithm 2 gNB Side Group-Level Scheduling

Require: User groups $\mathcal{G}$, Users precoder $\mathbf{W}$, DCIs
1: Initialize the scheduler discipline as round-robin
2: **for** each group $\mathcal{G}_g \in \mathcal{G}$ **do**
3: Allocate symbols and RBGs to $\mathcal{G}_g$ ▷ per-group basis
4: Stack the precoders $\{\mathbf{W}_j : j \in \mathcal{G}_g\}$
5: Compute the ZF precoder $\mathbf{W}_g^{\mathrm{ZF}}$
6: Generate the DCI for $\mathcal{G}_g$
7: **end for**

TABLE II
SIMULATION PARAMETERS

Parameter	Value
5G-LENA NR module	Release v4.1
Number of Monte Carlo runs	50
Simulation time	10 seconds
CBR traffic inter-packet interval time	$160\,\mu$s
CBR traffic packet size	1000 bytes
Channel model	3GPP SCM TR 38.901 UMi
Channel update period	50 ms
LoS/NLoS condition model	3GPP TR 38.901 UMi
Carrier frequency	3.5 GHz
Bandwidth	10 MHz
gNB transmit power	28 dBm
gNB and UE antennas	dual-polarized UPA 3GPP
MIMO codebook	Type-I and Type-II Codebooks
WB and SB update intervals	50 ms and 50 ms
gNB and UE number of panels	Single-Panel
gNB and UE antenna ports	$N_t = 32,\ N_r = 4$
gNB and UE antenna spacing	$d_h = 0.5\lambda,\ d_v = 0.5\lambda$
gNB and UE antenna height	10 and 1.5 meters
Numerology	0 (15 kHz subcarrier spacing)
Duplexing mode	TDD
TDD slot format	19
MCS Table	Table 2, TS 38.214 Table 5.1.3.1-2
UE processing delay (K1)	2 slots
RLC mode	UM
Scheduler Discipline	round-robin
Number of users K	$\{1, 2, 4, 8, 16\}$
δ_{thr}	0.03

A. Simulation Parameters & Scenarios

We use our ns-3 5G-LENA MU-MIMO implementation in Section IV to simulate a downlink scenario consisting of a single gNB and $K = \{1, 2, 4, 8, 16\}$ UEs. The carrier frequency is set to 3.5 GHz with a system bandwidth of 10 MHz. The channel is modeled using the 3GPP TR 38.901 Urban Micro (UMi) scenario, which captures realistic urban propagation characteristics, including path loss, shadowing, and small-scale fading, with both line-of-sight (LoS) and non-LoS (NLoS) conditions modeled according to the same specification. The channel realization is updated every 50 ms, and the gNB and UEs are both equipped with a single-panel dual-polarized UPA configured according to the 3GPP specifications, with $N_t = 32$ and $N_r = 4$ antenna ports, respectively. The gNB transmit power is set to 28 dBm, and the antenna heights are fixed at 10 m for the gNB and 1.5 m for the UEs. The UEs are randomly distributed within a 300 m radius of the gNB and move according to the random waypoint mobility model [19] at a constant speed of 3 km/h, changing direction every 5 m.

We assume a time-division duplexing (TDD) frame structure with numerology $\mu = 0$ (15 kHz subcarrier spacing). The slot format is configured so that the first and last symbols in a slot are dedicated to downlink (DL) and uplink (UL) control messages, while the rest are configured as flexible (F) symbols. The UE processing delay is set to two slots, which indicates that the scheduling decisions are made for two slots in the future, and the radio link control (RLC) layer operates in unacknowledged mode (UM). Traffic is modeled using a constant bit rate (CBR) source with a packet size of 1000 bytes and an inter-packet interval of 160 μs, ensuring a fully backlogged system.

The CSI is reported with a periodicity of 50 ms, and the maximum number of streams per transmission is set to $L_{\max} = 4$ (consistent with the existing ns-3 5G-LENA SU-MIMO implementation). Consequently, we limit the maximum number of simultaneously scheduled UEs to 4, where each UE is restricted to a single transmission layer per scheduling interval, and the CPOG compatibility threshold is set to $\delta_{thr} = 0.03$. We note that the compatibility threshold is chosen based on a grid search to optimize the interference management and the number of single-user groups. For a fair comparison, all benchmark MU-MIMO approaches are also configured to use a single transmission layer per UE. For the benchmark explicit CSI MU-MIMO methods, at each CSI reporting instance, the channel information of each UE is reported and stored at the gNB side. Each simulation is repeated over 50 independent Monte Carlo runs with different random seeds. The simulation parameters are summarized in Table II.

B. Performance Evaluation

We compare the performance of the proposed CPOG MU-MIMO against the MU-MIMO benchmarks in Section III-B. For codebook-based MU-MIMO, we compare our proposed CPOG user grouping to the k-means benchmark and consider both Type-I and Type-II codebooks in order to evaluate the impact of advanced codebook-based CSI feedback. The explicit CSI MU-MIMO benchmark schemes (NS-ZF, D-ZF, and k-means with explicit CSI) serve as a reference for the achievable performance with explicit (full) CSI, i.e., without the practical 3GPP constraint of implicit CSI feedback. Finally, we include TDMA as the baseline performance with no spatial multiplexing.

Let us first consider the aggregate throughput performance of the different MU-MIMO schemes with a different number of users in the network. Fig. 3 presents the distributions (over all Monte Carlo simulations) of the aggregate PHY throughput of correctly decoded packets for each MU-MIMO scheme as boxplots, for a network with $K = \{1, 2, 4, 8, 16\}$ UEs. Figs. 4 and 5 present the corresponding average co-scheduled users per symbol in a scheduling instance for $K = \{1, 2, 4, 8, 16\}$ and the SINR CDF of each scheme for $K = \{1, 2, 16\}$ UEs in the network, respectively. Fig. 3 confirms that for the $K = 1$ single user case, where no spatial user multiplexing is possible, all MU-MIMO schemes achieve a similar mean throughput of around 40–47 Mbit/s. The TDMA baseline without any spatial

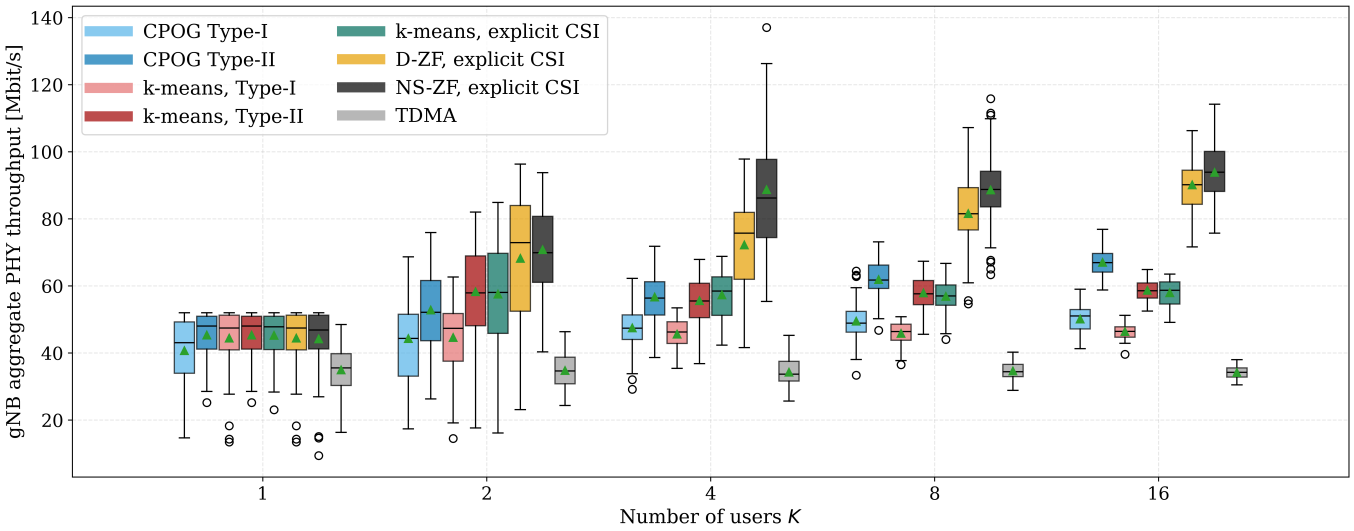


Fig. 3. Distribution of aggregate network throughput with $K = \{1, 2, 4, 8, 16\}$ users, for different MU-MIMO schemes (boxplot indicates lower-quartile/median/upper-quartile, and whiskers indicate minimum/maximum values up to 1.5 times the inter-quartile range; green triangle indicates mean).

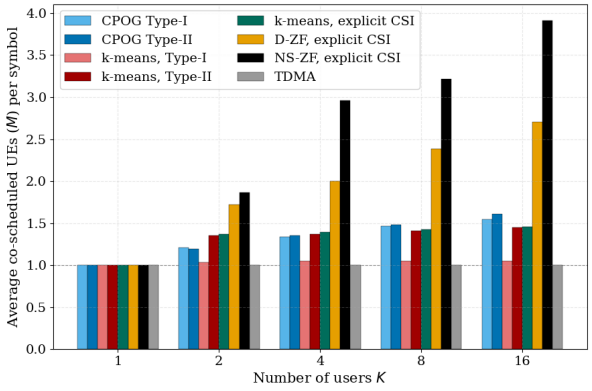


Fig. 4. Average number of co-scheduled users (M) per symbol for different MU-MIMO schemes for a network with $K = \{1, 2, 4, 8, 16\}$ users.

multiplexing achieves a mean aggregate throughput of around 35 Mbit/s regardless of K (which is slightly lower than the MU-MIMO single-user throughput due to differences in the selected precoder and post-SINR processing at the UE side). Fig. 5 shows that the single-user performance is consistent with the SINR distribution of all schemes, showing similar behavior when there is no co-scheduling and no inter-user interference, i.e., only one user is scheduled in each scheduling instance in Fig. 4.

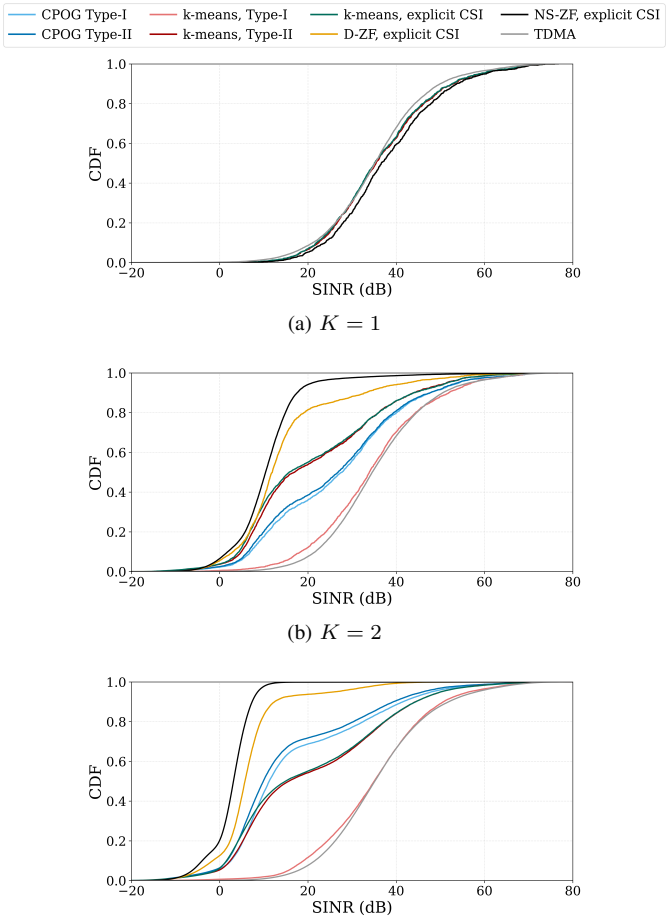
For the multi-user case, all MU-MIMO schemes achieve a higher throughput in Fig. 3 than the single-user case by grouping more users in Fig. 4, and correspondingly having a lower SINR in Fig. 5. We note that the "step" evident in the MU-MIMO SINR distributions in Figs. 5b and 5c is due to single-user groups formed by the MU-MIMO schemes when a user is not compatible with other users in the network (e.g., multi-user groups show a lower SINR than the single-user groups for the CPOG Type-II method in Fig. 5b). For $K \geq 2$, the NS-ZF baseline always performs best in Fig. 3, since it selects the optimal precoder and groups users based on the complete channel matrix, achieving, e.g., a mean throughput of around 90 Mbit/s with $K = 4$ UEs. The D-ZF method performs second-best (around 77.8 Mbit/s mean throughput with $K = 4$) since it uses explicit CSI but lacks the complete channel information available to the NS-ZF method and uses a greedy algorithm to form the user groups. This grouping behavior is indicated by the average number of co-scheduled users and the SINR distribution in Figs. 4 and 5, where NS-ZF and D-ZF form groups with more users during scheduling and achieve lower SINR than other methods. For instance, with $K = 16$, the NS-ZF method schedules on average around 4 users in a group, consistent with our ns-3 simulation limit of $L_{max} = 4$ spatial streams, and has the lowest SINR compared to the other MU-MIMO schemes due to inter-user interference inferred by the co-scheduled users on separate spatial streams. The D-ZF scheme comes second in terms of average co-scheduled users, scheduling on average 2.6 users in a network with $K = 16$. Additionally, Fig. 3 shows that the Mu-MIMO schemes exhibit higher variability in throughput compared to the single-user case. The larger throughput variation for the NS-ZF and D-ZF schemes (up to $\approx$ 137 Mbit/s at $K = 4$) compared to other MU-MIMO schemes indicates that explicit CSI MU-MIMO schemes, with less selective user grouping, are more sensitive to channel realization than the codebook-based schemes. OVerall, this reflects the trade-off of spatial multiplexing: MU-MIMO schemes achieve higher rates but have greater variability, since the instantaneous rate depends on the co-scheduled users and the resulting inter-user interference, while TDMA serves a single UE per symbol and

therefore delivers a stable but low rate.

Moreover, the throughput of the proposed CPOG scheme for both Type-I and Type-II codebooks in Fig. 3 increases monotonically with the number of users, indicating that CPOG can form feasible groups as the network scales. This is confirmed by Fig. 4, where the average number of co-scheduled users per symbol grows with the number of UEs in the network, reaching around 1.6 users for $K = 16$, showing that more compatible users are grouped together to be spatially multiplexed. The CPOG Type-II outperforms its counterpart, CPOG Type-I, as it can group the users more efficiently due to the richer codebook and construct a better precoding matrix. This behavior is also shown in Fig. 5: since the CPOG Type-II co-schedules a larger number of users, it achieves a lower SINR, and the "step" in the SINR distribution consequently shifts to higher percentiles for $K \geq 2$.

Importantly, Fig. 3 shows that among the considered codebook-based MU-MIMO schemes using the same Type I/II codebook, the proposed CPOG scheme achieves either a comparable (for $K \leq 4$) or higher throughput (for $K > 4$) than the k-means benchmark. This indicates that CPOG outperforms k-means by grouping spatially compatible users more efficiently in Fig. 4. Consequently, the SINR distribution in Fig. 5c shows that CPOG yields a higher percentile of users having lower SINR than its k-means counterpart due to co-scheduled users inter-user interference. We note that the k-means Type-I has an SINR comparable to the TDMA in Fig. 5 since it rarely groups the compatible users together due to the coarse Type-I codebook with limited feedback about the channel matrix. By contrast, even with the Type-I codebook, CPOG groups more users due to the precoder compatibility criterion. We also note that the "step" in the SINR distribution for the CPOG appears at higher percentiles than that for k-means schemes in Fig. 5c, which is the result of k-means more frequently forming single-user groups. Furthermore, Fig. 3 shows that the CPOG Type-II scheme outperforms the k-means explicit CSI MU-MIMO for $K > 4$ since it achieves a better spatial multiplexing efficiency (*cf.* Fig. 4) by grouping the users based on their Type-II precoders. By contrast, although the k-means explicit CSI scheme uses the full channel matrix for each user to form the groups, it cannot form spatially compatible groups, as two users in distinct clusters can remain incompatible. Consequently, the k-means Type-II performs close to k-means explicit CSI, since the Type-II codebook approximates the channel matrix's dominant direction. Overall, the results in Fig. 3 show that the explicit CSI MU-MIMO schemes achieve the highest aggregate throughput by co-scheduling more users, but have higher variability due to sensitivity to channel realizations. Among the codebook-based schemes with limited channel feedback, the proposed CPOG consistently matches or outperforms k-means schemes, including k-means with explicit CSI.

Finally, let us compare the proposed CPOG and the baseline MU-MIMO schemes (*cf.* Section III-B) in terms of computational complexity and the achievable throughput performance. Fig 6 shows the trade-off between the achievable average



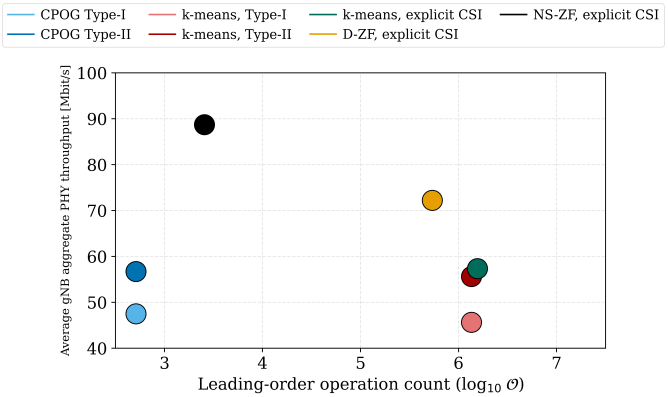
(a) $K = 1$

(b) $K = 2$

(c) $K = 16$

Fig. 5. SINR distribution over the network for different MU-MIMO schemes, with (a) $K = 1$, (b) $K = 2$, and (c) $K = 16$ users.

aggregate PHY throughput (*cf.* Fig. 3) and the complexity of the MU-MIMO schemes, in terms of leading-order operation count, at the gNB side, as stated in Table I with $K = 4$ and $K = 16$ users. Ideally, the perfect MU-MIMO scheme has low complexity and high throughput, i.e., being on the top left corner of the Fig. 6. For $K = 4$ and $K = 16$, the NS-ZF scheme achieves the highest throughput while having lower computational complexity than the D-ZF and k-means schemes. However, the explicit channel feedback required for the NS-ZF makes it impractical in 3GPP cellular deployments, especially as the number of UEs or antenna ports grows. The D-ZF method achieves the highest throughput gains as the number of UEs increases, about 18 Mbit/s from $K = 4$ to $K = 16$, but also incurs the largest complexity increase. Additionally, D-ZF's complexity exceeds that of k-means for $K = 16$ since increasing the number of users affects D-ZF more compared to the k-means scheme (as the K^2 term dominates the complexity in Table I). The CPOG has the lowest complexity and matches or exceeds the throughput of the k-means method for $K = 4$ and $K = 16$, respectively, demonstrating the efficacy of the proposed user grouping method. The complexity of CPOG grows with the number



(a) $K = 4$

(b) $K = 16$

Fig. 6. Scatterplot of average gNB aggregate PHY throughput vs. computational complexity for different MU-MIMO schemes with (a) $K = 4$, and (b) $K = 16$ users.

of UEs, following the same trend as D-ZF, since they both use a greedy algorithm to select user groups. Moreover, it is upper-bounded by the complexity of NS-ZF scheme, so it never exceeds this baseline complexity. Although CPOG and k-means Type-I/II have the same precoder complexity, their substantial complexity gap arises from differences in their user-selection approaches. Importantly, the CPOG Type-II scheme requires less channel feedback signaling than the explicit CSI methods to compute precoders and form user groups, making it the most attractive practical MU-MIMO scheme. Finally, the trend demonstrates the importance of codebook-based MU-MIMO user grouping methods to minimize both computational complexity and the channel feedback required to form compatible groups.

VI. Conclusions

We proposed CPOG, a novel codebook-based user grouping scheme for MU-MIMO that utilizes the Type-I/II codebooks to group spatially compatible users solely based on their precoders rather than full channel matrices through a greedy approach. In order to evaluate the system-level performance of CPOG in 5G-NR networks, we presented an open-source, full-stack framework of downlink MU-MIMO in the ns-3 5G-LENA simulator. Our ns-3 framework integrates the complete operation chain of CSI feedback generation, Type-II precoder selection at the UE, precoder-based user grouping, and scheduling at the gNB within a single simulation environment, enabling reproducible evaluations. Through extensive simulations, we benchmarked the performance of the proposed CPOG MU-MIMO against k-means codebook-based benchmark and explicit CSI-based user grouping methods like NS-ZF, D-ZF, and k-means. Overall, our results show the efficacy of CPOG as a practical codebook-based MU-MIMO scheme and highlight the importance of full-stack modeling for evaluating 5G MU-MIMO, where CSI feedback, user grouping, precoding, and scheduling decisions are tightly coupled. Future work will extend the framework to realize uplink MU-MIMO for 5G-NR systems.

References

[1] G. Femenias *et al.*, "Scalable cell-free massive mimo-based integrated sensing and communication," *IEEE Trans. Veh. Technol.*, vol. 74, no. 10, pp. 15643–15659, Oct. 2025.

[2] J. Zheng *et al.*, "Mobile cell-free massive MIMO: Challenges, solutions, and future directions," *IEEE Wireless Communications*, vol. 31, no. 3, pp. 140–147, June 2024.

[3] D. Gesbert *et al.*, "Shifting the MIMO paradigm," *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 36–46, Sept. 2007.

[4] 3rd Generation Partnership Project (3GPP), "NR; Physical layer procedures for data," Release 18, Version 18.6.0 3GPP TS 38.214, Technical Specification Group Radio Access Network (TSG RAN), Apr. 2025.

[5] E. Castaneda *et al.*, "An overview on resource allocation techniques for multi-user MIMO systems," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 239–284, Firstquarter 2016.

[6] L.-U. Choi and R. D. Murch, "A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach," *IEEE Trans. Wireless Commun.*, vol. 3, no. 1, pp. 20–24, 2004.

[7] G. Caire *et al.*, "On the achievable throughput of a multiantenna gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, July 2003.

[8] L.-N. Tran *et al.*, "Iterative precoder design and user scheduling for block-diagonalized systems," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3726–3739, July 2012.

[9] P. Lu *et al.*, "Sum-rate analysis of multiuser MIMO system with zero-forcing transmit beamforming," in *Proc. in IEEE ICC*, Bejing, 2008.

[10] R.-F. Trifan *et al.*, "K-means MU-MIMO user clustering for optimized precoding performance," in *Proc. in IEEE VTC2019-Spring*, Kuala Lumpur, 2019.

[11] [Online]. Available: https://www.nsnam.org/.

[12] B. Bojović *et al.*, "3GPP-compliant single-user MIMO model for high-fidelity mobile network simulations," *Computer Networks*, vol. 256, p. 110912, 2025.

[13] G. Ferreira *et al.*, "5G-LENA alternative rank and precoding matrix indicator selection techniques," in *Proc. in the 2025 International Conference on ns-3*, New York, 2025.

[14] F. Gómez-Cuba *et al.*, "Hybrid beamforming in 5G mmWave networks: A full-stack perspective," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 1288–1303, Feb. 2022.

[15] X. Fu *et al.*, "A tutorial on downlink precoder selection strategies for 3GPP MIMO codebooks," *IEEE Access*, vol. 11, pp. 138897–138922, Dec. 2023.

[16] L.-N. Tran *et al.*, "Multiuser diversity for successive zero-forcing dirty paper coding: Greedy scheduling algorithms and asymptotic performance analysis," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3411–3416, June 2010.

[17] B. H. Wang *et al.*, "Maximum volume criterion for user selection of multiuser MIMO downlink with multiantenna users and block diagonalization beamforming," in *Proc. in IEEE Antennas and Propagation Society International Symposium*, Toronto, 2010.

[18] [Online]. Available: https://github.com/mcc-rwth.

[19] X. Lin *et al.*, "Towards understanding the fundamentals of mobility in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, pp. 1686–1698, April 2013.