electronics

MDPI

Article

Spectral Efficiency of Precoded 5G-NR in Single and Multi-User Scenarios under Imperfect Channel Knowledge: A Comprehensive Guide for Implementation

David Alejandro Urquiza Villalonga [1,*], Hatem OdetAlla [2], M. Julia Fernández-Getino García [1] and Adam Flizikowski [2]

[1] Department of Signal Theory and Communications, Universidad Carlos III de Madrid, 28911 Madrid, Spain
[2] IS-Wireless, 05-500 Piaseczno, Poland
* Correspondence: daurquiza@tsc.uc3m.es

Abstract: Digital precoding techniques have been widely applied in multiple-input multiple-output (MIMO) systems to enhance spectral efficiency (SE) which is crucial in 5G New Radio (NR). Therefore, the 3rd Generation Partnership Project (3GPP) has developed codebook-based MIMO precoding strategies to achieve a good trade-off between performance, complexity, and signal overhead. This paper aims to evaluate the performance bounds in SE achieved by the 5G-NR precoding matrices in single-user (SU) and multi-user (MU) MIMO systems, namely Type I and Type II, respectively. The implementation of these codebooks is covered providing a comprehensive guide with a detailed analysis. The performance of the 5G-NR precoder is compared with theoretical precoding techniques such as singular value decomposition (SVD) and block-diagonalization to quantify the margin of improvement of the standardized methods. Several configurations of antenna arrays, number of antenna ports, and parallel data streams are considered for simulations. Moreover, the effect of channel estimation errors on the system performance is analyzed in both SU and MU-MIMO cases. For a realistic framework, the SE values are obtained for a practical deployment based on a clustered delay line (CDL) channel model. These results provide valuable insights for system designers about the implementation and performance of the 5G-NR precoding matrices.

Keywords: 5G New Radio; 3GPP; multi-antenna beamforming; digital precoding; multiple-input multiple-output (MIMO)

Citation: Urquiza Villalonga, D.A.; OdetAlla, H.; Fernández-Getino García, M.J.; Flizikowski, A. Spectral Efficiency of Precoded 5G-NR in Single and Multi-User Scenarios under Imperfect Channel Knowledge: A Comprehensive Guide for Implementation. *Electronics* **2022**, *11*, 4237. https://doi.org/10.3390/electronics11244237

Academic Editors: Adão Silva, Daniel Castanheira and Rui Dinis

Received: 30 November 2022
Accepted: 15 December 2022
Published: 19 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Multiple-input multiple-output (MIMO) systems have been widely used with beamforming techniques to enhance spectral efficiency (SE). The main idea of multi-antenna beamforming is to concentrate the radiation pattern energy toward a desired angular direction where the target user is located. Therefore, the signal-to-noise ratio (SNR) at the receiver side is increased while interference with near users is avoided. The use of multiple antennas in a massive MIMO (mMIMO) approach provides a higher array gain and a narrower beam. As the number of transmit antennas grows, the beam becomes more directive, improving both the SE and energy efficiency (EE). MIMO systems provide diversity which is essential to improve link reliability and robustness against fading channels. Furthermore, spatial multiplexing is also enabled with digital beamforming, commonly known as precoding. This arises as a key solution to achieve the expected values of data rate in 5G and beyond 5G (B5G) networks. Data symbols could be spatially multiplexed through multiple parallel layers using the same time and frequency resources and thus improving the SE. This is achieved by properly designing the precoding and combining matrices at the transmitter and receiver side, respectively [1–4].

Several digital precoding techniques have been developed to obtain a multi-layer transmission to a single-user MIMO (SU-MIMO) or multi-user MIMO (MU-MIMO). The

most commonly used linear precoding techniques reported in the scientific literature are singular value decomposition (SVD), zero forcing (ZF), maximum ratio transmission (MRT), maximum ratio combining (MRC), and minimum mean squared error (MMSE) [3,5–7]. In [5], the capacity of SU-MIMO Gaussian channels is studied. This work proves that SVD is the optimal precoding scheme for SU-MIMO systems. Then, authors in [6] explain the paradigm shift of MIMO systems when a multi-user scenario is considered. The advantages and challenges of different precoding techniques such as ZF and MMSE for MU-MIMO systems are described. In [7], the concept of massive MIMO is introduced to increase the array gain. The conjugate beamforming for downlink and matched filter decoding for uplink have been illustrated. In [3], a summary of the main precoding techniques for SU-MIMO and MU-MIMO is provided. Recently, the use of multi-antenna beamforming techniques in combination with reconfigurable intelligent surfaces (RIS) has been investigated to improve the SE and energy-efficiency of the network [8–11]. Furthermore, novel beamforming methods have been proposed for satellite systems [12,13].

The above techniques have been proposed for single-carrier and multi-carrier waveforms such as orthogonal frequency division multiplexing (OFDM). However, these methods are based on the perfect knowledge of all the coefficients of the channel state information (CSI) matrix at both the transmitter and receiver sides. This is unfeasible in real systems due to the large signal overhead to communicate the estimated CSI matrix between the transmitter and receiver. Therefore, several efforts have been made to obtain solutions that achieve the best trade-off between performance and signal overhead. With this aim, codebook-based strategies are implemented to select from a set of predefined precoding matrices the one that achieves the best system performance. To reduce feedback overhead, the selection of the precoder at the transmitter side is only based on a quantified version of the estimated CSI matrix by the receiver. This is a well-studied topic in the state of the art [14–21] and it has already been applied in the standardization of the communication systems [2,3].

1.1. Related Works

Codebook-based MIMO precoding techniques have been proposed in LTE (Long Term Evolution) through ten transmission modes to provide high diversity, array gain, and spatial multiplexing [22,23]. However, the increasing demands of data rate and the necessity to provide communication services to a much denser network with heterogeneous technologies make advanced MIMO precoding even more crucial in 5G New Radio (NR). Therefore, the 3rd Generation Partnership Project (3GPP) has developed and improved the existing LTE advanced multi-antenna techniques since Release 15 to enable codebook-based precoded transmission mode in 5G [24,25]. The structure of the vectors that integrate the codebooks used in 5G-NR resembles the discrete Fourier transform (DFT); thus, they are typically denoted as DFT-codebooks. They support multi-layer transmissions with low signal overhead and complexity [2].

The 5G-NR precoding matrix implementation for both the downlink and uplink is described in the 3GPP technical specifications (TSs) 38-211 [24] and 38-214 [25]. The downlink precoding design is much more complex than the uplink counterpart because the NR base station (gNB) is equipped with a higher number of antennas than the user equipment (UE). Therefore, more sophisticated processing is required to coherently combine the signals into multiple paths. Two types of downlink precoding, named Type I and Type II, have been proposed since Release 15 with different spatial resolution properties. Type I codebook is mainly applied in SU-MIMO and hence it is focused on reporting information of the strongest channel path to focus the energy on the target UE. On the contrary, Type II is addressed to support communication to multiple UEs in an MU-MIMO approach. Therefore, more accurate channel estimation is required to avoid interference between different UEs at the cost of large-signal feedback overhead. This is the main issue in Type II that is improved in Release 16 with the incorporation of an enhanced precoding strategy that reduces feedback overhead and provides finer frequency granularity [4,25]. Several

works have been proposed to reduce feedback overhead in Type II codebook based on a multi-stage quantization approach [26] and a DFT-compression [27,28]. In Release 17, the CSI feedback overhead is further reduced by exploiting reciprocity in Frequency Division Duplex (FDD) operation [29]. NR standardization process is continuously evolving and it is foreseen that more advanced precoding techniques will be part of the forthcoming releases as a key feature for 5G/B5G networks.

1.2. Motivation

Nevertheless, despite the efforts in standardization to include codebook-based MIMO techniques in the last releases, the information provided in the above TSs is very limited to support a clear understanding of the precoding matrices design. For this reason, several books have compiled the information on 5G-NR standardization and have exposed it more clearly [2–4,29–32]. In [2,3], a description of the CSI acquisition, the main physical reference signals, and the general procedure of the design of the precoding matrices are covered. A study about the implementation of Type I single-panel, Type I multi-panel, and Type II is given in [30,31]. The improvements introduced in Release 16 and 17 are described in [29,32]. However, a detailed analysis matching the theory with the implementation stages in the TSs is missing.

Although several studies have been proposed to evaluate the 5G-NR performance under a realistic 3GPP channel model, the new sets of features in recent releases are described mainly focusing on the services, technology, and use cases [33,34]. In [35], authors propose a system-level simulator for evaluating the 3GPP performance in the 5G use cases according to the requirements defined by the International Telecommunication Union (ITU). However, in all these previous works, an analysis of the precoding matrix design and the SE bounds that can be attained in SU-MIMO or MU-MIMO scenarios is missing. This study is required to provide valuable insights for system designers about the implementation and performance of a practical deployment of a 5G-NR network. Therefore, to fill this gap in the scientific literature, an initial approach consisting of a detailed description of the 5G-NR precoding matrix design and its performance in SU-MIMO systems is addressed in [36]. However, this paper is limited to the assumption of perfect channel knowledge and the effect of channel estimation errors is not taken into account. Also, simulation results were obtained with a tapped delay line (TDL) model that does not properly represent the spatial characteristic of the channel. Moreover, an MU-MIMO scenario is not considered, which restricts its applicability in dense networks. Therefore, motivated by these preliminary results, the present work provides a complete and comprehensive guide for the implementation of the 5G-NR precoding matrices for both SU and MU-MIMO scenarios, even considering channel estimation impairments and channel models aware of spatial characteristics. To the best of our knowledge, the proposed analysis is missing in the literature.

1.3. Contribution

This paper studies the performance bounds in terms of SE achieved by the 5G-NR precoding matrices in SU-MIMO and MU-MIMO systems. The main contributions of this paper are summarized as follows:

- A detailed description of the 5G-NR precoding matrices implementation covering both Type I and Type II codebook is provided. Each implementation step outlined in the standard is complemented with the corresponding theoretical explanation to improve its understanding, thus providing a comprehensive guide.
- The SE bounds of the 5G-NR precoding matrices are first obtained for a SU-MIMO scenario by using the Type I codebook. These suboptimal precoding matrices are compared with the optimal singular value decomposition (SVD) solution [5] in order to quantify the margin of improvement that could be attained in future precoding designs. It has been found that an increase in the number of parallel layers does

not always imply a higher SE and the best choice has been determined in several configurations.Also, the impact on performance of linear or planar arrays is addressed.
- Then, these results are extended to an MU-MIMO system based on the Type II codebook design. The performance achieved with the standardized solution is compared with the theoretical block-diagonalization ZF method described in [37,38]. The saturation effect experienced in SE for high SNR values due to interference is exhibited. It is demonstrated how the usage of Type II codebook is limited to users with low spatial correlation.
- The effect of imperfect channel estimation on system performance is numerically evaluated in both cases, single-user and multi-user.
- Moreover, the simulations are performed for several configurations of antenna arrays, number of antenna ports, and parallel data streams. To guarantee a realistic scenario, a clustered delay line (CDL) channel model defined in the 3GPP technical report (TR) 38-901 is considered.

Numerical results validate the use of these advanced multi-antenna techniques to reach the demanding SE in 5G/B5G networks. It must be stated that the novelty of the paper relies on a new performance analysis of the SE bounds achieved by the 5G-NR precoding matrices. Additionally, it can be observed the gap between the codebook based standardized solutions and the optimal ones. Furthermore, this article serves as a guide for the practical implementation of precoding strategies which can be very useful for system designers. The obtained results provide relevant insights for the deployment of 5G and beyond networks.

1.4. Paper Organization

This paper is organized as follows. The system model is presented in Section 2. The description of the computation of the precoding matrix compliant with the 5G-NR standard is provided in Section 3 for SU-MIMO systems based on Type I single-panel and multi-panel codebooks. Then, Type II codebook design for MU-MIMO systems is detailed in Section 4. Numerical results are presented in Section 5 considering a realistic deployment of a 5G-NR scenario by using the CDL channel model defined in TR 38-901. Conclusions are finally drawn in Section 6.

Notation: boldface lower-case letters are used for vectors, while boldface upper-case letters are used for matrices. $\mathbf{A}^T$, $\mathbf{A}^H$, and $\mathbf{A}^{-1}$ represent the transpose, conjugate transpose, and inverse of matrix $\mathbf{A}$, respectively. $\mathbb{E}(\mathbf{A})$ stands for the expected value of $\mathbf{A}$. $[\mathbf{A}]_d$ represents the dth column vector of matrix $\mathbf{A}$ while $\mathbf{A}(m,n)$ is the (m,n)th element. $A \otimes B$ represents the Kronecker product on matrices A and B. $\binom{n}{k}$ denotes the binomial coefficient. $\|\cdot\|$, $\|\cdot\|_F$ are used to denote the Euclidean and Frobenius norm, respectively. $\mathbb{C}^{N\times M}$ is the space of complex $N \times M$ matrices. $\mathcal{CN}(\mu, \sigma^2)$ is the complex Gaussian distribution with mean μ and variance σ^2. $\mathbf{I}_d$ is the $d \times d$ identity matrix.

2. System Model

In this section, a general MU-MIMO setting composed of K UEs is considered as shown in Figure 1. At each symbol period, the gNB sends N_s data streams spatially multiplexed to K UEs. An OFDM transmission with N_{sc} subcarriers in the frequency domain is assumed. The complex symbols transmitted at the jth subcarrier are denoted as $\mathbf{s}_j = \left[\mathbf{s}_j^{[1]}, ..., \mathbf{s}_j^{[K]}\right]^T \in \mathbb{C}^{N_s K\times 1}, \forall j \in \{1, 2, ..., N_{sc}\}, \forall k \in \{1, 2, ..., K\}$. In this equation, $\mathbf{s}_j^{[k]} \in \mathbb{C}^{N_s\times 1}$ represents the N_s data streams transmitted to the kth UE. It is assumed that $\mathbb{E}\left[\mathbf{s}_j^{[k]}\left(\mathbf{s}_j^{[k]}\right)^H\right] = \mathbf{I}_{N_s}$. The gNB and each UE are equipped with N_t and N_r antennas, respectively.

The OFDM symbols are precoded in the gNB to avoid interference between the UEs. The precoding matrix is designed according to the 5G-NR standard. To avoid large overhead, contiguous subcarriers are grouped in a subband since they are likely to

experience a comparable channel effect. The number of subcarriers in each subband is given by $N_f = N_{sc}^{RB} N_{PRB}^{SB}$, where N_{sc}^{RB} is the number of subcarriers in each physical resource block (PRB) and N_{PRB}^{SB} is the number of PRB in each subband. Let $\mathbb{A} = \{\mathcal{A}_1, ..., \mathcal{A}_n, ..., \mathcal{A}_{N_{SB}}\}$ denote the set of subbands, where $\mathcal{A}_n$ contains the subcarriers that are in the nth subband. The total number of subbands N_{SB}, given by the cardinality of the set $N_{SB} = |\mathbb{A}|$, is defined according to the TS 38-214 [25]. The subcarriers inside the nth subband (i.e., $\forall j \in \mathcal{A}_n$) are precoded with the same matrix $\mathbf{W}_n$ computed according to the 5G-NR standard. For simplicity, it is assumed that the number of physical transmit antennas is the same as that of virtual antenna ports. Therefore, the precoding matrix is given by $\mathbf{W}_n = \left[\mathbf{W}_n^{[1]}, ..., \mathbf{W}_n^{[K]}\right] \in \mathbb{C}^{N_t \times N_s K}$, where $\mathbf{W}_n^{[k]} \in \mathbb{C}^{N_t \times N_s}$ is the precoder related to the kth UE.

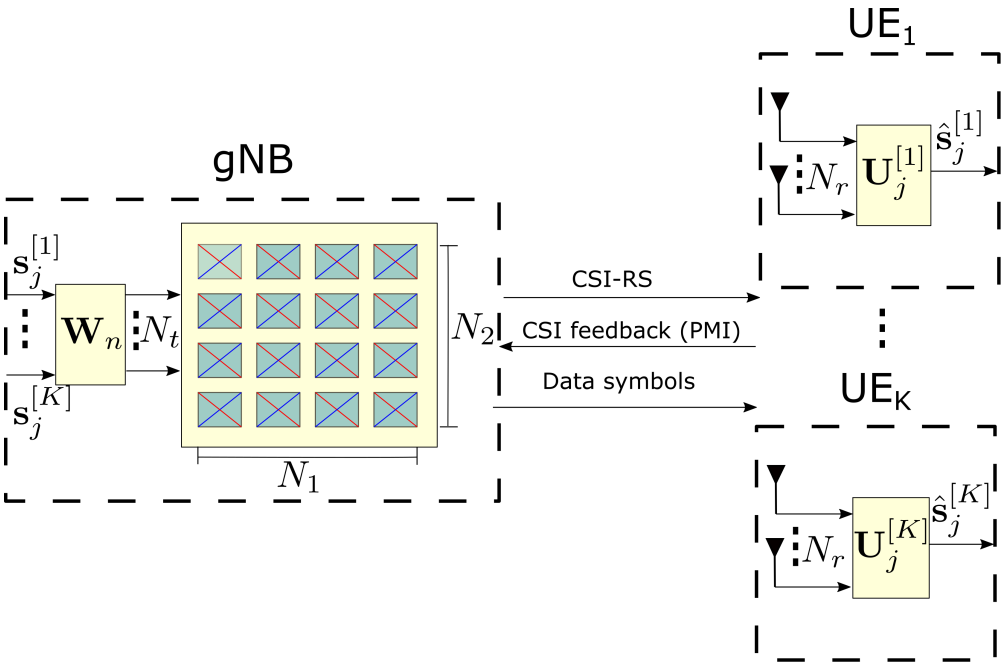

Figure 1. Block diagram of an MU-MIMO downlink transmission in 5G-NR composed of several stages. First, the gNB sends the CSI-RS to each UE. Based on this signal, the channel is estimated at each UE and a CSI is fedback, including the PMI. With this information, the gNB designs the precoding matrix. Finally, the precoded data symbols are transmitted to each UE avoiding interference.

A codebook-based beamforming is considered, according to 5G-NR specifications. Therefore, the precoding matrix is selected from a set of predefined codebooks based on the knowledge that the gNB can acquire from the CSI. Therefore, a codebook-based CSI acquisition procedure is applied [2]. As illustrated in Figure 1, the channel is estimated at the UE by measuring known reference signals transmitted from the gNB called CSI reference signal (CSI-RS). A quantized version of the estimated CSI is reported from the UE to the gNB including information about the precoding matrix indicator (PMI) and rank indicator (RI), which identify the best precoding matrix according to the UE measurements. This information is used by the gNB as a recommendation to select the precoding matrix. CSI feedback has been designed in the 5G-NR standard to cover both Type I and Type II codebooks, for SU-MIMO and MU-MIMO, respectively [25,39].

The received signal at the kth UE side is multiplied by the combining matrix $\mathbf{U}_j^{[k]} \in \mathbb{C}^{N_r \times N_s}$ to maximize the signal to interference and noise ratio (SINR). With these considerations, the received symbol vector is expressed as [6]

$$\hat{\mathbf{s}}_j^{[k]} = \left(\mathbf{U}_j^{[k]}\right)^H \mathbf{H}_j^{[k]} \mathbf{W}_n^{[k]} \mathbf{s}_j^{[k]} + \left(\mathbf{U}_j^{[k]}\right)^H \mathbf{H}_j^{[k]} \sum_{i=1, i \neq k}^{K} \mathbf{W}_n^{[i]} \mathbf{s}_j^{[i]} + \left(\mathbf{U}_j^{[k]}\right)^H \mathbf{n}_j^{[k]}, \forall j \in \mathcal{A}_n, \quad (1)$$

where $\mathbf{H}_j^{[k]} \in \mathbb{C}^{N_r \times N_t}$ represents the channel matrix at the jth subcarrier between the gNB and the kth UE. Furthermore, $\mathbf{n}_j^{[k]} \in \mathbb{C}^{N_r \times 1}$ is the circularly symmetric additive white Gaussian noise vector (AWGN) at the kth UE $\left(\mathbf{n}_j^{[k]} \sim \mathcal{CN}\left(0, \sigma_j^{2[k]} \mathbf{I}_{N_r}\right)\right)$. The first term in (1) represents the desired signal while the second term is the interference produced by the multiplexed UEs that share the same PRBs. The SE at the jth subcarrier and kth UE is computed as

$$R_j^{[k]} = \sum_{d=1}^{N_s} \log_2\left(1 + SINR_j^{[kd]}\right), \tag{2}$$

where $SINR_j^{[kd]}$ denotes the SINR at the jth subcarrier, kth UE, and dth data stream, $\forall d \in \{1, 2, ..., N_s\}$. For a given precoding and combining matrix, the SINR can be written as [40]

$$SINR_j^{[kd]} = \frac{\left|\left[\mathbf{U}_j^{[k]}\right]_d^H \mathbf{H}_j^{[k]} \left[\mathbf{W}_n^{[k]}\right]_d\right|^2}{\left[\mathbf{U}_j^{[k]}\right]_d^H \mathbf{Q}_j^{[kd]} \left[\mathbf{U}_j^{[k]}\right]_d}, \forall j \in \mathcal{A}_n. \tag{3}$$

It must be pointed out that the denominator in (3) includes the interference terms and also the noise variance, expressed through the covariance matrix $\mathbf{Q}_j^{[kd]} \in \mathbb{C}^{N_r \times N_r}$. This is detailed in the following equation where the interference plus noise covariance matrix $\mathbf{Q}_j^{[kd]}$ is defined as

$$\mathbf{Q}_j^{[kd]} = \sum_{i=1}^{K} \mathbf{H}_j^{[i]} \mathbf{W}_n^{[i]} \left(\mathbf{W}_n^{[i]}\right)^H \left(\mathbf{H}_j^{[i]}\right)^H - \mathbf{H}_j^{[k]} \left[\mathbf{W}_n^{[k]}\right]_d \left[\mathbf{W}_n^{[k]}\right]_d^H \left(\mathbf{H}_j^{[k]}\right)^H + \sigma_j^{2[k]} \mathbf{I}_{N_r}. \tag{4}$$

The major goal of this paper is to evaluate the SE bounds achieved by the precoding matrices designed within the 5G-NR specifications. However, the standard does not specify the combining filter design at the receiver side for downlink data transmission since it is vendor specific based on their own research. In this work, the MMSE combining method is applied to maximize the SINR based on the precoding matrix previously computed at the gNB side [3,40]. With this strategy, the highest SE is attained. Nevertheless, the design of the combining matrix is limited by the knowledge that the UE can acquire from the channel. It must be noted that in a multi-user scenario, each UE is only able to estimate its channel to the gNB supported by the CSI-RS. However, a full representation of the channels with the remaining users can not be obtained. Therefore, according to this restriction, the combining matrix is designed as

$$\left[\mathbf{U}_j^{[k]}\right]_d = \frac{\left(\tilde{\mathbf{Q}}_j^{[kd]}\right)^{-1} \mathbf{H}_j^{[k]} \left[\mathbf{W}_n^{[k]}\right]_d}{\left\|\left(\tilde{\mathbf{Q}}_j^{[kd]}\right)^{-1} \mathbf{H}_j^{[k]} \left[\mathbf{W}_n^{[k]}\right]_d\right\|}, \forall j, \forall k, \forall d, \tag{5}$$

where the covariance matrix $\tilde{\mathbf{Q}}_j^{[kd]}$ only considers the channel between the gNB and the kth UE as follows

$$\tilde{\mathbf{Q}}_j^{[kd]} = \mathbf{H}_j^{[k]} \mathbf{W}_n^{[k]} \left(\mathbf{W}_n^{[k]}\right)^H \left(\mathbf{H}_j^{[k]}\right)^H - \mathbf{H}_j^{[k]} \left[\mathbf{W}_n^{[k]}\right]_d \left[\mathbf{W}_n^{[k]}\right]_d^H \left(\mathbf{H}_j^{[k]}\right)^H + \sigma_j^{2[k]} \mathbf{I}_{N_r}. \tag{6}$$

The subsequent sections describe the design of the precoding matrix according to the 5G-NR standard for SU-MIMO and MU-MIMO systems.

3. Precoding Matrix Design Based on Type I 5G-NR for SU-MIMO

The design of the precoding matrix according to the Type I codebook for SU-MIMO is described in this section. This is a particular case of the system model described in Section 2 assuming $K = 1$. The main goal is to transmit the signal over the N_s strongest propagation

paths of the channel matrix to the target UE. Inter-layer interference is avoided with the precoding matrix designed at the gNB based on the PMI reported from the UE. Table 1 defines the supported parameters for Type I.

Table 1. Supported parameters by Type I downlink codebook-based transmission.

Type	Maximum Number of Layers	Number of Antenna Ports	Number of Antenna Panels	Supported Configurations of (N_1, N_2) and (O_1, O_2)
Type I Single-Panel	8	2, 4, 8, 12, 16, 24, 32	1	Table 5.2.2.2.1-2 TS-38.214
Type I Multi-panel	4	8, 16, 32	2, 4	Table 5.2.2.2.2-1 TS-38.214

According to TS 38-214 [25], the gNB is configured with a maximum number of 32 virtual antenna ports for the downlink transmission. Without loss of generality, the virtual antenna ports are assumed to be equal to the physical transmit antennas in this paper. Then $N_t = 2N_g N_1 N_2$, where N_g is the number of antenna panels, N_1 and N_2 are the dual-polarized (DP) antenna elements in the horizontal and vertical dimension, respectively, as illustrated in Figure 1. Note that the product by 2 indicates the use of DP antennas. A single or multi-panel antenna is supported where each panel can be designed following a uniform linear (ULA) or planar array (UPA). Up to $N_s = 8$ and $N_s = 4$ layers transmission is available for single-panel and multi-panel, respectively. The maximum number of layers N_s depends on the number of antenna ports such that $N_s \leq \min(N_t, N_r)$. Since more users are commonly located in the azimuth plane, then a higher number of antennas is available in the horizontal dimension to provide a higher spatial resolution. A maximum number of 16 antenna elements in a single-panel array is supported with a configuration of $N_1 = 16$, $N_2 = 1$ for ULA or $N1 = 8$, $N_2 = 2$ for UPA [31].

The main objective in Type I is to select N_s precoding vectors from a predefined DFT-codebook set that corresponds to the strongest propagation paths. The precoding vector for the dth layer in the nth subband is designed as

$$[\mathbf{W}_n]_d = \mathbf{W}_1^{[d]} \mathbf{w}_{2n}^{[d])}, \tag{7}$$

where $\mathbf{W}_1^{[d]} \in \mathbb{C}^{N_t \times 2}$ and $\mathbf{w}_{2n}^{[d])} \in \mathbb{C}^{2 \times 1}$ are related to wideband and subband parameters, respectively. Note that the index k is omitted in (7) since a single user is considered in this section. The wideband matrix $\mathbf{W}_1^{[d]}$ is defined as follows

$$\mathbf{W}_1^{[d]} = \begin{bmatrix} \mathbf{b}^{[d]} & \mathbf{0} \\ \mathbf{0} & \mathbf{b}^{[d]} \end{bmatrix}, \tag{8}$$

where $\mathbf{b}^{[d]} \in \mathbb{C}^{N_1 N_2 \times 1}$ indicates the direction of the strongest beam in the dth layer. This factor is independent of the frequency and polarization since it only depends on large-scale properties of the channel [31]. Note that since orthogonal polarization is assumed, then the same beam is transmitted over the two sets of DP antennas without causing interference. Consequently, $\mathbf{W}_1^{[d]}$ has a diagonal block structure. For an UPA, the dth beam $\mathbf{b}^{[d]}(l_d, m_d)$ is obtained from a DFT-codebook set as a function of the horizontal and vertical indexes l_d and m_d, respectively, as follows [2,30]

$$\mathbf{b}^{[d]}(l_d, m_d) = \mathbf{w}_{ULA}(l_d) \otimes \mathbf{w}_{ULA}(m_d), \tag{9}$$

where $\mathbf{w}_{ULA}(l_d)$ and $\mathbf{w}_{ULA}(m_d)$ are the DFT-based ULA precoders in the horizontal and vertical direction, respectively, given by

$$\begin{aligned} \mathbf{w}_{ULA}(l_d) = \quad & \frac{1}{\sqrt{N_1}}\left[1 \quad e^{j2\pi\frac{l_d}{O_1N_1}}, \ \ldots, e^{j2\pi(N_1-1)\frac{l_d}{O_1N_1}}\right]^T, \\ & l_d \in \{0,1,...,O_1N_1-1\}, \\ \mathbf{w}_{ULA}(m_d) = \quad & \frac{1}{\sqrt{N_2}}\left[1 \quad e^{j2\pi\frac{m_d}{O_2N_2}}, \ \ldots, e^{j2\pi(N_2-1)\frac{m_d}{O_2N_2}}\right]^T, \\ & m_d \in \{0,1,...,O_2N_2-1\}. \end{aligned} \tag{10}$$

In the case of ULA, above expressions are simplified considering index $m_d = 0$ and $N_2 = 1$. The terms O_1 and O_2 define the oversampling factors in the horizontal and vertical dimensions, respectively, to obtain a finer spatial granularity. This increases the probability of always being able to steer the beam in a direction close to the UE and hence, the beamforming gain is improved. The beam selection step is obtained by properly selecting the horizontal (l_d) and vertical (m_d) indexes from the predefined DFT-codebook set with dimension $N_1N_2O_1O_2$. Note that one beam is selected per layer, and the N_s beams (i.e., $\mathbf{b}^{[d]}, \forall d \in \{1,...,N_s\}$) form an orthogonal basis to avoid inter-layer interference. The orthogonality can be obtained in the horizontal dimension between beams spaced by an integer factor of O_1, in the vertical dimension between beams spaced by an integer factor of O_2, or in both dimensions at the same time. The values of O_1 and O_2 are not configurable parameters in NR standard. For UPA, $O_1 = O_2 = 4$ and for ULA, $O_1 = 4, O_2 = 1$ (see Table 5.2.2.2.1-2 in TS 38-214 [25] for details).

On the other hand, $\mathbf{w}_{2n}^{[d]}$ denotes the inter-polarization co-phasing factor which is a frequency-selective subband parameter defined as

$$\mathbf{w}_{2n}^{[d]} = \begin{bmatrix} 1 \\ e^{j\,\phi_n^{[d]}} \end{bmatrix}, \quad \phi_n^{[d]} \in \left\{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\right\}. \tag{11}$$

Based on the structure of the precoding matrix in (7), the PMI is divided into the set of indexes $\mathbf{i}_1$ and $\mathbf{i}_2^n$ that are used for the gNB to compute $\mathbf{W}_1^{[d]}$ and $\mathbf{w}_{2n}^{[d]}$, respectively, according to the Tables 5.2.2.2.1-5 to 5.2.2.2.1-12 for single-panel and Tables 5.2.2.2.2-3 to 5.2.2.2.2-5 for multi-panel in TS 38-214 [25]. Each table contains the steps to design the precoding matrix according to the codebook type (i.e., single-panel or multi-panel), the specific antenna array configuration of the gNB given by N_g, N_1, N_2, and the number of parallel layers N_s [25]. Furthermore, there are two different codebook modes that define the structure of $\mathbf{W}_1^{[d]}$. Note that $\mathbf{i}_2^n$ needs to be reported for each subband but $\mathbf{i}_1$ is the same for the entire bandwidth. Thus, feedback overhead is reduced. The next sections describe the design of the precoding matrix for a single and multi-panel array.

3.1. Type I Single-Panel

Type I single-panel codebook is designed assuming that all antennas are grouped into a single panel (i.e., $N_g = 1$) with a DP ULA or UPA structure. The wideband PMI index is defined as $\mathbf{i}_1 = [i_{11}, i_{12}, i_{13}]$, where $i_{11} \in \{0,1,\ldots,N_1O_1-1\}$ and $i_{12} \in \{0,1,\ldots,N_2O_2-1\}$ are mapped into l_1 and m_1, respectively. For the case of ULA, m_d is always fixed to 0 and i_{12} is not reported. Then, the index $i_{13} \in \{0,1,2,3\}$ is used to obtain the relative position of the higher layer beams with respect to $\mathbf{b}^{[1]}$ in case of $2 \leq N_s \leq 4$. In the other cases (i.e., $5 \leq N_s \leq 8$), the phase offset of the higher layer beams with respect to $\mathbf{b}^{[1]}$ is fixed. Finally, the inter-polarization co-phasing factor is obtained with $i_2^n \in \{0,1,2,3\}$.

To provide a better understanding of the beam selection procedure, Figure 2 illustrates the radiation pattern of all the beams that can be obtained with a 5G-NR UPA antenna panel considering $N_1 = N_2 = 2$. Each of these beams represents the radiation pattern to a particular angular direction in the azimuth and zenith plane. Then, Figure 3 depicts a simpler model of this radiation pattern to understand the composition of the DFT grid of beams. Note that this is a representation of the main radiation pattern of each beam without

considering the secondary lobes which is a very effective way of illustration used also in [30,31]. The grid is composed of a total of $A_c = 64$ beams. The number of orthogonal DFT beams without the oversampling factors is given by $N_1 N_2 = 4$ and they are represented by the red-filled circles. The remaining 60 circles represent the oversampled or rotated beams.

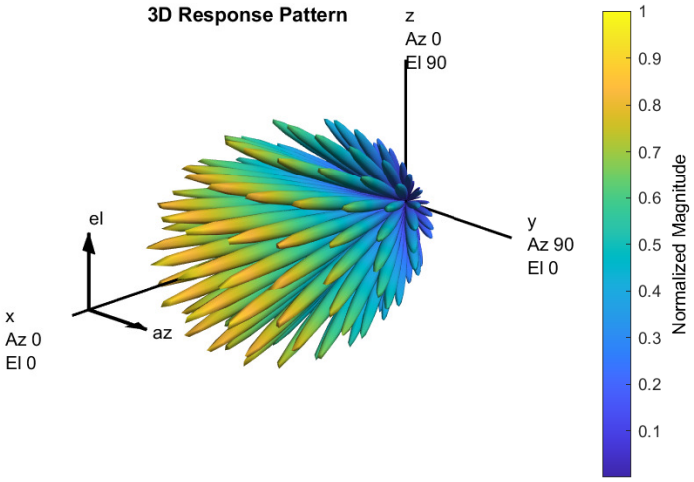

Figure 2. Radiation pattern of all the beams obtained with a 5G-NR antenna array designed as a DP

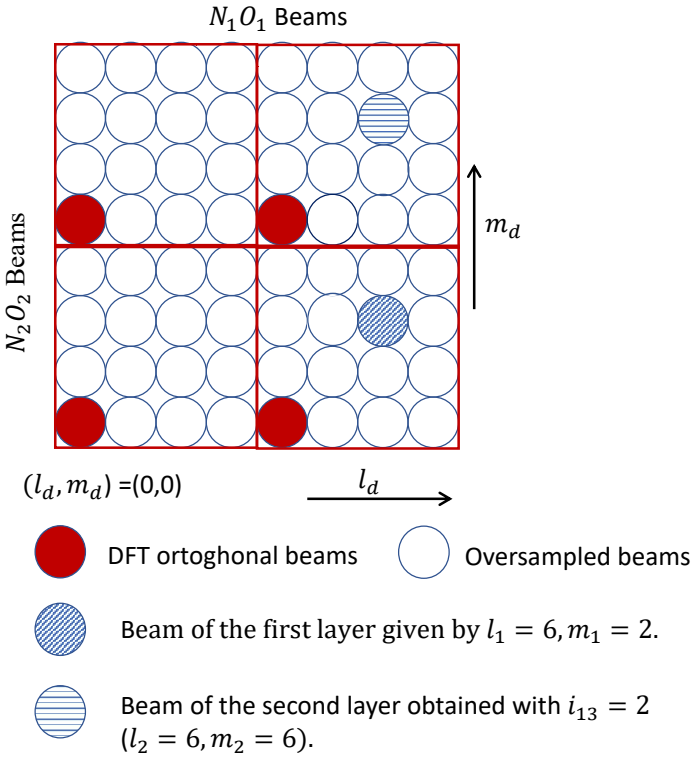

Figure 3. Grid of beams of a single-panel UPA with $N_1 = 2, N_2 = 2$, and $O_1 = O_2 = 4$. It is assumed that $N_s = 2, i_{11} = 6, i_{12} = 2$, and $i_{13} = 2$.

The wideband precoding matrix $\mathbf{W}_1^{[d]}$ selects a beam per layer from the predefined DFT grid. In this example, it is assumed that $N_s = 2, i_{11} = 6, i_{12} = 2$, and $i_{13} = 2$. The striped circles illustrate the $N_s = 2$ orthogonal beams that are chosen according to the indexes l_d and m_d. The beam of the first layer is obtained with $l_1 = i_{11} = 6$, and $m_1 = i_{12} = 2$. Then, assuming that $i_{13} = 2$, the beam for the second layer is given by $l_2 = 6$, and $m_2 = 6$ (see Tables 5.2.2.2.1-3 and 5.2.2.2.1-6 in TS 38-214 [25] for details). It can be noted that the beams

are orthogonal in the vertical dimension. Finally, these two beams are combined with the co-phasing factors selected per subband. Note that the UE only needs to report the coefficients and co-phasing factor of the beam in the first layer. Then, the gNB is capable of computing the beams and the co-phasing factors for the remaining layers. This strategy reduces feedback overhead guaranteeing an efficient transmission.

Commonly, the same matrix $\mathbf{W}_1^{[d]}$ is used for the entire bandwidth. However, codebook mode 2 enables a special case for $N_s = 1$ or $N_s = 2$ where $\mathbf{W}_1^{[d]}$ may vary for different subbands. In this case, the selected beam also depends on the frequency-selective coefficient i_2^n. This represents a more flexible design at the cost of requiring more bits for reporting i_2^n. Therefore, it is only implemented for a low number of layers (i.e., $N_s = 1$ or $N_s = 2$) to ensure that the overhead is not too high. There is another special design of the matrix $\mathbf{W}_1^{[d]}$ enabled for ($N_t \geq 16$) when $N_s = 3$ or $N_s = 4$. The main goal is to prevent the beam in the horizontal dimension from becoming narrower as the number of antenna ports increases, which can be harmful to some propagation models. Therefore, the available values of i_{11} are reduced to $i_{11} \in \left\{0, 1, \ldots, \frac{N_1 O_1}{2} - 1\right\}$ to widen the beam [25,31].

3.2. Type I Multi-Panel

The precoding matrix design for multi-panel antenna arrays is just an extension of Type I single-panel. In this case, the wideband PMI index is defined as $\mathbf{i}_1 = [i_{11}, i_{12}, i_{13}, i_{141}, i_{142}, i_{143}]$, and the frequency-selective index is $\mathbf{i}_2^n = \left[i_{20}^n, i_{21}^n, i_{22}^n\right]$. The main problem with using several antenna panels is that the distance between two consecutive antennas of different panels does not have to be equal to the spacing between two consecutive antennas in the same panel. Therefore, the assumption of uniform arrays is no longer fulfilled. To solve this problem, the beamforming design is based on the assumption that all antenna panels are identical. Each panel is composed of N_1 antenna elements in the horizontal direction and N_2 in the vertical direction. Then, the idea is to design the precoding matrix $\mathbf{W}_{SP_n} \mathbb{C}^{2N_1 N_2 \times N_s}$ for a single-panel. Finally, this matrix is repeated in the remaining panels but adding a phase offset $e^{j\alpha_p}$, where $p \in \{1, ..., N_g - 1\}$, to compensate the non-uniform distance between panels.

Currently, the precoding matrix can be obtained according to two different codebook modes supporting the spatial multiplexing up to 4 parallel layers. The codebook mode 1 supports both $N_g = 2$ and $N_g = 4$. The co-phasing between each panel is obtained using only wideband coefficients. On the contrary, in the case of codebook mode 2, the co-phasing is configured by applying a combination of wideband and subband parameters. This mode is only supported for two antenna panels. In general, the precoding vector for the nth subband and dth layer can be written as

$$[\mathbf{W}_n]_d = \begin{bmatrix} \mathbf{w}_{SP_n}^{[d]} \\ \ldots \\ e^{j\alpha_p} \mathbf{w}_{SP_n}^{[d]} \\ \ldots \\ e^{j\alpha_{N_g-1}} \mathbf{w}_{SP_n}^{[d]} \end{bmatrix}, \tag{12}$$

where

$$\mathbf{w}_{SP_n}^{[d]} = \begin{bmatrix} \mathbf{b}^{[d]} & \mathbf{0} \\ \mathbf{0} & \mathbf{b}^{[d]} \end{bmatrix} \begin{bmatrix} 1 \\ e^{j\phi_n^{[d]}} \end{bmatrix}, \quad \phi_n^{[d]} \in \left\{0, \tfrac{\pi}{2}, \pi, \tfrac{3\pi}{2}\right\}. \tag{13}$$

The phase offset $e^{j\alpha_p}$ is computed according to the wideband sub-index $\mathbf{i}_{14}$ in codebook mode 1, such that $\alpha_p = \frac{\pi}{4} + \frac{\pi i_{14p}}{2}$, where $i_{14p} \in \{0, 1, 2, 3\}$. The coefficients for the higher layer are also adjusted based on the subindex i_{13} to obtain an orthogonal precoding matrix.

4. Precoding Matrix Design Based on Type II 5G-NR for MU-MIMO

In this section, the design of the precoding matrices based on the Type II codebook is described. The main idea is to support communication to multiple UEs that share the same PRBs in MU-MIMO systems. In this case, a more sophisticated precoding design is required to cancel the interference in the spatial domain between the UEs at the cost of higher feedback overhead. In Release 15, only 2 data streams per UE are supported. However, enhanced precoding strategies are proposed in Release 16 to provide a finer frequency granularity supporting up to 4 layers and reducing signal overhead. Type II is only designed for single-panel antenna arrays with the configurations defined in Table 5.2.2.2.1-2 [25].

The main characteristic of Type II is that it applies a linear combination of L orthogonal beams per layer instead of just selecting the strongest one as in Type I. This strategy provides a more accurate representation of the dominant eigenvectors of the channel to obtain a precoding matrix that avoids interference between the UEs. Similar to Type I, this design can be represented as a two-step procedure for the kth user and dth layer denoted by $\left[\mathbf{W}_n^{[k]}\right]_d = \mathbf{W}_1^{[kd]}\mathbf{w}_{2n}^{[kd]}$. However, in the case of Type II codebook, the wideband matrix $\mathbf{W}_1^{[kd]} \in \mathbb{C}^{N_t \times 2L}$ represents an orthogonal basis composed of L beams per polarization and layer, where $L \in \{2, 3, 4\}$. Then, a linear combination of the chosen beams is performed by using $\mathbf{w}_{2n}^{[kd]} \in \mathbb{C}^{2L \times 1}$ which is comprised of wideband and subband complex coefficients. With this design, the desired signal is coherently combined at the receiver side, while the interference signal from co-located UEs is canceled improving the SINR [2,30].

The wideband matrix for the kth user and dth layer is obtained as

$$\mathbf{W}_1^{[kd]} = \begin{bmatrix} \mathbf{B}^{[kd]} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}^{[kd]} \end{bmatrix}, \tag{14}$$

where $\mathbf{B}^{[kd]} = \left[\mathbf{b}_0^{[kd]}, ..., \mathbf{b}_{L-1}^{[kd]}\right] \in \mathbb{C}^{N_1 N_2 \times L}$ is composed of L beams obtained from a predefined DFT-codebook. Note that $\mathbf{b}_l^{[kd]}$, $\forall l = \{0, 1, ..., L-1\}$, is defined in (9) for the dth layer and kth UE based on a given horizontal ($l_d^{[k]}$) and vertical ($m_d^{[k]}$) indexes. The beams in $\mathbf{B}^{[kd]}$ form an orthogonal basis to avoid inter-layer interference. Similar to Type I, the orthogonality can be obtained in the horizontal, vertical, or both dimensions at the same time.

The PMI index set to obtain the wideband and subband matrices is defined as $\mathbf{i}_1 = [i_{11}, i_{12}, i_{13}, i_{14}]$, and $\mathbf{i}_2^n = \left[i_{21}^n, i_{22}^n\right]$, respectively. The dimension of each subindex is given by $i_{11} \in \mathbb{R}^{2\times 1}$, $i_{12} \in \mathbb{R}^{1\times 1}$, $i_{13} \in \mathbb{R}^{N_s \times 1}$, $i_{14} \in \mathbb{R}^{2LN_s \times 1}$, $i_{21}^n \in \mathbb{R}^{2LN_s \times 1}$, and $i_{22}^n \in \mathbb{R}^{2LN_s \times 1}$. Note that a higher overhead is required than in Type I. The computation of the wideband matrix $\mathbf{W}_1$ is performed according to the $\mathbf{i}_1$ index vector. The first problem of this design arises due to the long feedback required to identify the L beams for each layer and polarization. The main solution to reduce overhead is to first select L DFT orthogonal beams based only on the information reported in the $i_{12} \in \left\{0, 1, ..., \binom{N_1 N_2}{L} - 1\right\}$ subindex. The algorithm proposed in section 5.2.2.2.3 of TS 38-214 [25] selects the L beams based on a binomial coefficients computation. Then, to increase the granularity, these DFT beams are oversampled in the horizontal and vertical dimensions by the subindex $\mathbf{i}_{11}$. The same rotated factors are used for all the L DFT beams to reduce feedback overhead.

Then, these selected L beams are linearly combined with different amplitude and phase weights for each layer and polarization according to $\mathbf{w}_{2n}^{[kd]}$ as follows

$$\mathbf{w}_{2n}^{[kd]} = \mathbf{P}_{WB}^{[kd]}\mathbf{P}_{SB_n}^{[kd]}\mathbf{C}_n^{[kd]}. \tag{15}$$

where $\mathbf{P}_{WB}^{[kd]} \in \mathbb{R}^{2L\times 1}$ and $\mathbf{P}_{SB_n}^{[kd]} \in \mathbb{R}^{2L\times 1}$ represent the wideband and subband amplitude coefficient, respectively, for the dth layer and kth UE. The wideband amplitude parameter $\mathbf{P}_{WB}^{[kd]}$ is computed from the subindex vector $\mathbf{i}_{14}$ which is quantized with 3 bits. Therefore,

each element of $\mathbf{P}_{WB}^{[kd]}$ takes values from the set $\left\{0, \sqrt{1/64}, \sqrt{1/32}, \sqrt{1/16}, \sqrt{1/8}, \sqrt{1/4}, \sqrt{1/2}, 1\right\}$. The subband amplitude coefficients $\mathbf{P}_{SB_n}^{[kd]}$ are selected from the set $\left\{1, \sqrt{0.5}\right\}$ according to the subindex $\mathbf{i}_{22}^n$ that uses a single-bit quantization to reduce overhead.

On the other hand, $\mathbf{C}_n^{[kd]}$ denotes the inter-polarization co-phasing vector. Similarly to Type I, each value of $\mathbf{C}_n^{[kd]}$ is given by the complex exponential $e^{j\phi_n^{[kd]}}$. The phase values $\phi_n^{[kd]}$ are obtained from an N_{PSK}-phase-shift-keying (PSK) alphabet, where $N_{PSK} \in \{4, 8\}$, according to the subindex vector $\mathbf{i}_{21}^n$.

In Release 16, the Type II codebook is enhanced to support a higher number of parallel layers (i.e., $N_s = 4$) and a more accurate representation of the channel eigenvectors. In this sense, the amplitude and phase coefficients are obtained with a higher number of quantization bits. Furthermore, the frequency granularity is improved by the introduction of the "frequency domain (FD) units" concept. Consequently, the precoding matrices are designed per FD instead of per subband as in Release 15. Since the number of FD units can be twice the number of subbands, the frequency granularity can be improved by a factor of 2. However, this demands a higher CSI-RS feedback overhead. To overcome this problem a DFT-based compression technique has been proposed. For this purpose, the wideband $\mathbf{W}_1$ and linear combination $\mathbf{W}_2$ matrices reuse the same design concept of Type II codebook in Release 15 but they are providing the results per FD units. Moreover, an additional matrix is defined to perform the frequency compression $\mathbf{W}_f$ of certain DFT basis vectors from the total of FD units. Furthermore, to keep the CSI overhead controllable, either a one step or two steps feedback scheme is used based on the number of FD units. In Release 17, the angles and delays reciprocity in FDD operation are exploited to reduce computation complexity at the UE side. In addition, enhancements of multiple transmission and reception point (mTRP) are addressed [26,29,32].

5. Simulation Results

In this section, the SE bounds achieved by the 5G-NR precoding matrices are evaluated for SU-MIMO and MU-MIMO scenarios. Simulation results are obtained in MATLAB by using a CDL channel model with a delay spread of $300 \cdot 10^{-9}$ s, and a maximum Doppler shift of 50 Hz. A delay profile CDL-C defined according to TR 38-901 is assumed [38]. A different offset in the angles of departure in the azimuth (AoD) and zenith (ZoD) dimensions is added to each UE to simulate a deployment of K users randomly distributed with uniform probability within a circular coverage area. It is assumed that the gNB is located at a height of 30 m in the center of the circle. Then, the UEs are spread in the azimuth dimension between $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ and within a radio between 50 and 500 m from the gNB. The simulation results are averaged over 50 time slots and 100 random distributions of the UEs. The typical 5G-NR numerology is used for simulation as illustrated in Table 2.

Without loss of generality, the transmission power of the gNB is normalized to $\|\mathbf{W}_n\|_F^2 = 1$. This work is mainly focused on obtaining the SE bounds of the 5G-NR precoding matrices assuming a downlink data transmission. Therefore, the power control mechanism for different UEs is not considered. At each experiment, the PMI is obtained at the UE side by evaluating all the precoding matrices that can be computed according to the 5G-NR standard for a given antenna array configuration and a value of N_s. Finally, the PMI corresponding to the matrix that achieves the highest SINR is selected. For simplicity, codebook mode 1 is considered for the simulations.

Table 2. Simulation parameters.

Parameters	Value
Size of the bandwidth part	N_{BWP} = 52 PRB
Number of subcarriers per PRB	N_{sc}^{RB} = 12
Number of subcarriers of the OFDM grid	N_{sc} = 624
Subcarrier Spacing	Δ_f = 15 kHz
Subband Size	N_{PRB}^{SB} = 4 PRB
Number of subbands	N_{SB} = 13
Symbols per slot	N_{symb}^{slot} = 14
Slots per frame	N_{slot}^{frame} = 10
Number of frames	5
CSI-RS period	4 slots

5.1. SE Bounds Achieved by Type I Codebook

In this subsection, the performance bounds of the Type I codebook are evaluated for SU-MIMO systems. The SE is obtained averaging over all the subcarriers as $R = \frac{1}{N_{sc}} \sum_{j=1}^{N_{sc}} R_j$, where R_j is the SE per subcarrier defined in (2). Figures 4–6 illustrate the average SE for a range of SNR between −5 and 15 dB and assuming different antenna configurations at the gNB. The number of receive and transmit antennas is set to $N_r = 4$, and $N_t = 8$ (i.e., 8 CSI-RS antenna ports), respectively. Therefore, the maximum number of parallel layers allowed is $N_s = 4$. Figures 4 and 5 are obtained for a single-panel antenna array with $N_1 = 4$, $N_2 = 1$ (i.e., ULA), and $N_1 = 2$, $N_2 = 2$ (i.e., UPA), respectively. Figure 6 depicts the performance achieved by a multi-panel antenna array with $N_g = 2, N_1 = 2, N_2 = 1$.

The results obtained with the 5G-NR design are compared with the optimal solution given by the SVD of the channel matrix which serves as a benchmark [5]. It can be noted that for all the configurations shown, the performance of the SVD is better than 5G-NR solution at the cost of perfect knowledge of the channel matrix at both the gNB and UE sides. This is not a practical assumption in real scenarios as it would require an unfeasible overhead. On the contrary, the 5G-NR solution is designed to achieve a trade-off between performance, low overhead, and complexity. Therefore, since the gNB computes the precoding matrix based on a predefined codebook set and using the PMI index transmitted from the UE, the performance is degraded in comparison to the optimal SVD solution. However, the overhead is significantly reduced because only a few bits are required to transmit the PMI. In order to illustrate the complexity and overhead advantages of 5G-NR over SVD, a practical example will be shown. According to TS 38-212 [39] and TS 38-214 [25], in the case of single-panel ULA with $N_1 = 4$ and $N_s = 4$, the PMI requires 6 bits to report the wideband parameters and a single bit per subband to feedback the frequency selective coefficients. Since $N_{SB} = 13$, a total of 19 bits are used to compute the precoding matrix. On the contrary, for the same gNB configuration, the SVD needs to report an 8×4 matrix with complex coefficients from the UE to the gNB. Assuming an 8-bit quantization, a total of 6656 bits must be fed back. This amount of bits to report the CSI from the UE to the gNB is not affordable in practical applications because it leads to several drawbacks. First, the probability that the CSI will be correctly decoded by the gNB is lower than with smaller report sizes. Therefore, packet segmentation at the UE side might be needed. This means that the report is fed back in different scheduling periods which will introduce more latency and complexity. Moreover, packet retransmission could be a necessity and the delay in the network will be increased. It can be noted that a higher number of parallel layers does not always translate into a higher SE. The value of N_s that achieves the highest performance depends on the channel matrix, antenna configuration, and the value of SNR. Therefore, the UE reports to the gNB the RI of the precoding matrix that corresponds to the value of

N_s that attains the highest SE. In general, for low SNR values, it is better to concentrate the energy in a single very directional path to the UE. In this example, the values of N_s that achieve the highest performance averaging over all the SNR range, for the single-panel ULA, UPA, and multi-panel ULA are $N_s = 3$, $N_s = 1$, and $N_s = 2$, respectively. According to these figures, the best performance is obtained with the ULA configuration. The reason for this is that, in this case, the UPA layout has few antennas in each dimension, which does not properly exploit the spatial characteristics of the channel.

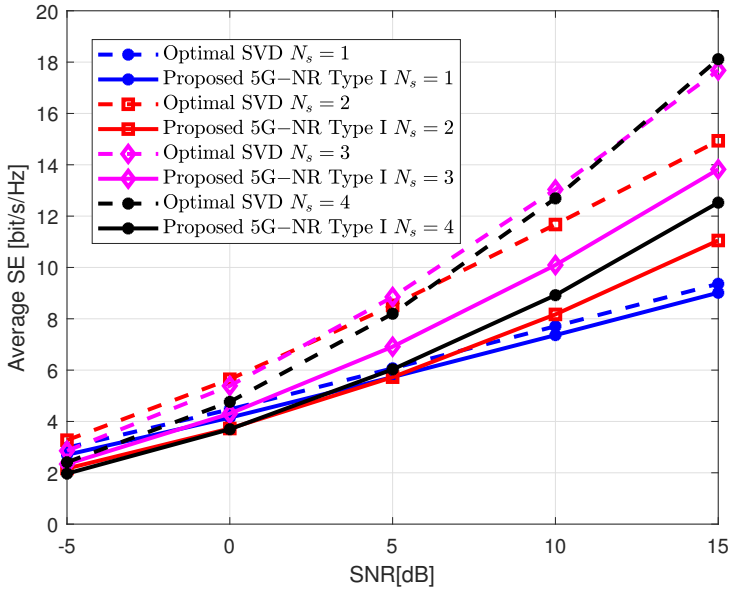

Figure 4. Average SE vs. SNR for SU-MIMO systems assuming $N_t = 8, N_r = 4$, and different values of N_s. The gNB is implemented assuming a single-panel ULA with four DP polarized antenna elements in the horizontal dimension ($N_1 = 4$ and $N_2 = 1$).

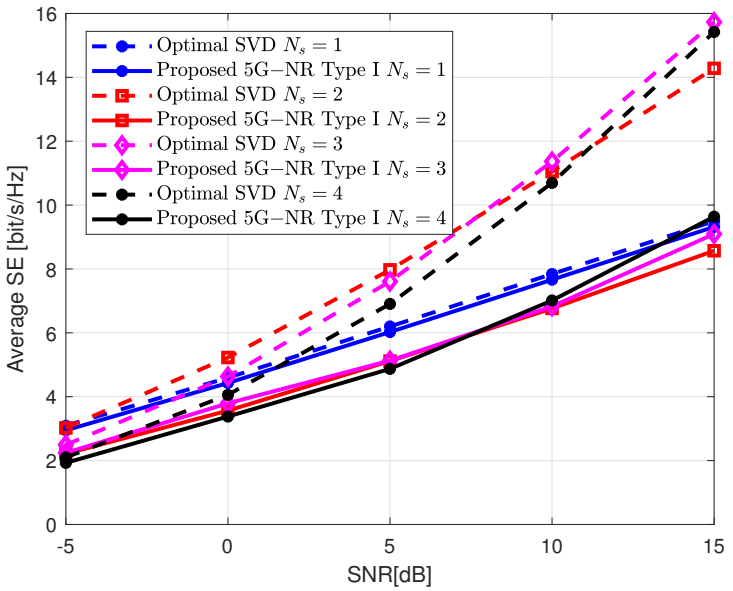

Figure 5. Average SE vs. SNR for SU-MIMO systems assuming $N_t = 8, N_r = 4$, and different values of N_s. The gNB is implemented assuming a single-panel UPA with two DP polarized antenna elements in the horizontal and vertical dimensions ($N_1 = 2$ and $N_2 = 2$).

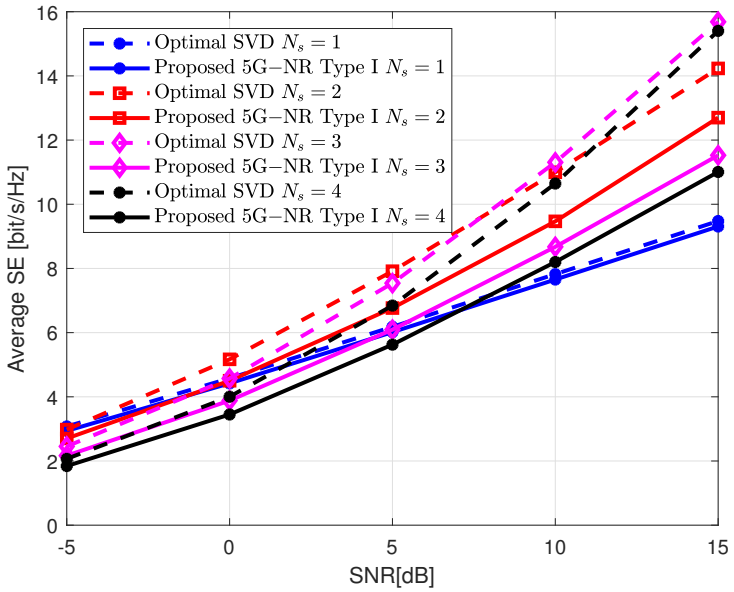

Figure 6. Average SE vs. SNR for SU-MIMO systems assuming $N_t = 8, N_r = 4$, and different values of N_s. The gNB is implemented assuming a multi-panel ULA with $N_g = 2, N_1 = 2$, and $N_2 = 1$.

Figure 7 shows the robustness of the 5G-NR precoding matrices in the presence of channel estimation errors. This result is obtained by adding an error matrix $\mathbf{E}_j \in \mathbb{C}^{N_r \times N_t}$ to the true channel $\mathbf{H}_j$ as $\hat{\mathbf{H}}_j = \mathbf{H}_j + \mathbf{E}_j$. Each element of $\mathbf{E}_j$ follows a Gaussian distribution with zero mean and variance σ_e^2. The performance loss is evaluated for two different error variances, $\sigma_e^2 = 10^{-3}$ and $\sigma_e^2 = 10^{-2}$, which are common values for channel estimation algorithms. The gNB is configured following the ULA layout used in Figure 4. According to the results of Figure 4, the number of parallel layers is now fixed to $N_s = 3$. It can be noted that the channel estimation error affects the 5G-NR precoding performance in the same proportion as the optimal SVD solution. The performance degradation is negligible for $\sigma_e^2 = 10^{-3}$. In the case of $\sigma_e^2 = 10^{-2}$, the performance loss is only noticeable for high values of SNR. This is due to the fact that the effect of noise is more relevant than the channel estimation errors for low SNR.

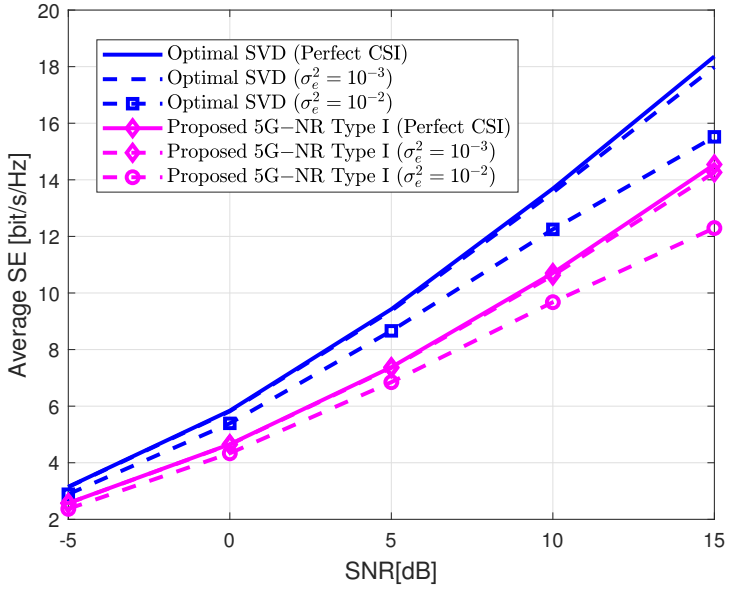

Figure 7. Average SE vs. SNR for SU-MIMO systems assuming $N_t = 8, N_r = 4$, and $N_s = 3$. The gNB is implemented assuming a single-panel ULA with four DP polarized antenna elements in the horizontal dimension ($N_1 = 4$ and $N_2 = 1$).

Figure 8 depicts the average SE for a higher number of transmit antennas given by $N_t = 32$ (i.e., 32 CSI-RS antenna ports). The performance is evaluated for several values of N_s. Similar to the previous figures, three types of different antenna configurations are considered, including both single and multi-panel arrays. A single-panel ULA and UPA are considered with $N_1 = 16, N_2 = 1$, and $N_1 = 4$, $N_2 = 4$, respectively. Furthermore, a multi-panel is considered with $N_g = 2, N_1 = 8$, and $N_2 = 1$. This figure shows the SE gain obtained with respect to Figures 4–6 when increasing from 8 to 32 antenna ports. The performance loss due to channel estimation error is also evaluated considering $\sigma_e^2 = 10^{-3}$. Similar to Figure 7, the degradation in SE is only slightly noticeable for SNR values higher than 10 dB. Regarding the overhead between linear and planar arrays, ULA demands lower overhead because the subindex i_{12} is not reported. For this particular example, the single-panel ULA requires a PMI with a total of 20 bits while single-panel UPA needs 22 bits. It must be pointed out that the multi-panel ULA also needs to report the subindex i_{141} and hence, 22 bits of overhead are used as in the case of single-panel UPA [25,39]. However, the computational complexity is higher in UPA because a DFT precoder is computed in both dimensions. Then, the Kronecker product is applied in (9) to obtain the final wideband precoding matrix. On the contrary, these expressions are simplified in ULA since only a DFT precoder in one dimension is required. Although the complexity and overhead are higher in UPA, its performance is superior since the beams can be steered in horizontal and vertical directions. It must be noted that the highest SE in Figure 8 is obtained with the UPA configuration because the number of antennas in both dimensions becomes higher and hence, the spatial properties of the channel are better exploited.

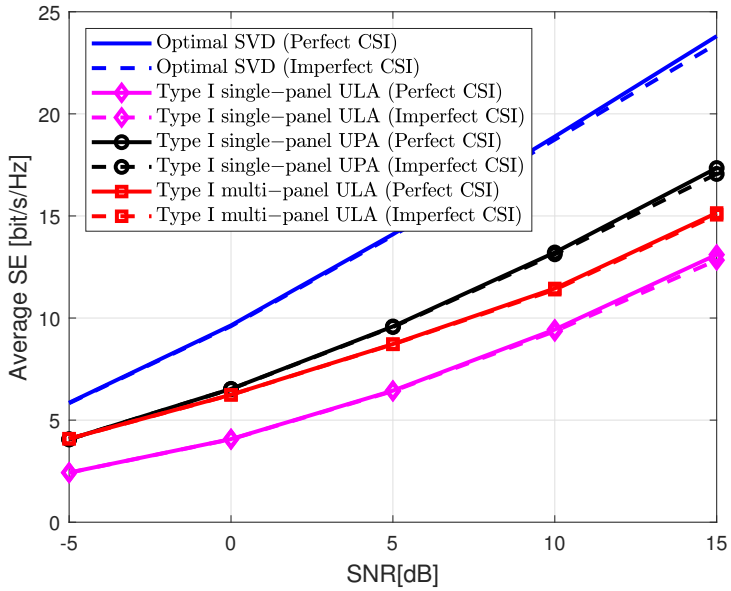

Figure 8. Average SE vs. SNR for SU-MIMO systems assuming $N_t = 32, N_r = 4$, and $N_s = 3$. The gNB is implemented with different antenna configurations. Imperfect CSI is also assumed with $\sigma_e^2 = 10^{-3}$. All the curves in this figure labeled as Type I refer to the proposed 5G-NR case.

5.2. SE Bounds Achieved by Type II Codebook

In this subsection, the performance obtained with the Type II codebook for MU-MIMO scenarios is illustrated. The total SE of the network is measured as $\sum_{k=1}^{K} R^{[k]}$, which is known as sum-SE. The sum-SE achieved by the 5G-NR precoding matrices is compared with the theoretical block-diagonalization (BD) method described in [37]. This method is designed to remove the interference signals and maximize the SINR in (3) based on the knowledge of the channel matrix of all UEs. A perfect and imperfect CSI with $\sigma_e^2 = 10^{-3}$ are assumed. In the case of 5G-NR, two beams per layer and polarization (i.e., $L = 2$) and a QPSK alphabet (i.e., $N_{PSK} = 4$) are considered.

Figure 9 illustrates the SE bounds assuming $N_t = 32, N_r = 4$, and $K = 2$. The gNB is implemented following a single-panel ULA with $N_1 = 16$, and $N_2 = 1$. Figure 10 shows the performance for the same scenario but considering a single-panel UPA with $N_1 = 4$, and $N_2 = 4$. In both cases, the SE of the 5G-NR is saturated for high SNR values. This is because interference is the limiting factor that avoids performance enhancement. It should be noted that the overhead in Type II is much higher than in Type I. For this particular example, assuming $N_s = 1$ and the single-panel ULA configuration, 22 bits are required to report the wideband coefficients while 9 bits per subband are needed for the frequency selective parameters. Therefore, a total of 137 bits per UE is required to report the PMI. In the case of UPA, two extra bits are required for the wideband parameters which leads to a total of 139 bits per UE. Similar to the Type I codebook, the effect of the channel estimation errors with $\sigma_e^2 = 10^{-3}$ hardly influences the SE results.

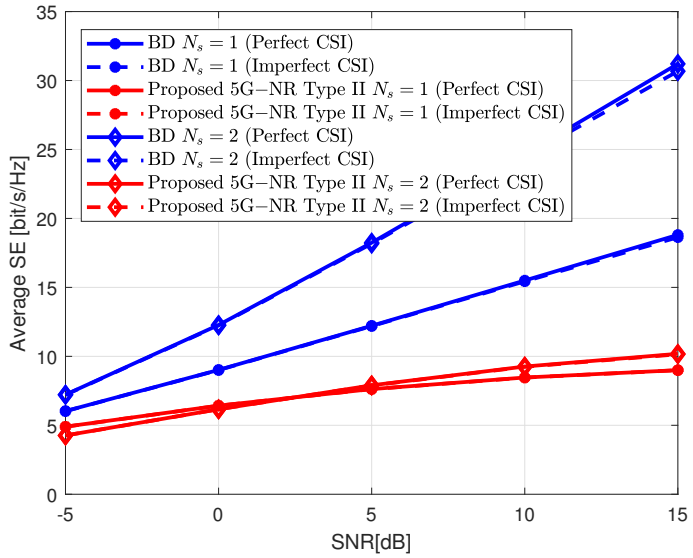

Figure 9. Average SE vs. SNR for MU-MIMO systems assuming $K = 2$, $N_t = 32, N_r = 4$, and different values of N_s. The gNB is implemented assuming an ULA single-panel with $N_1 = 16$, and $N_2 = 1$. Imperfect CSI is also assumed with $\sigma_e^2 = 10^{-3}$.

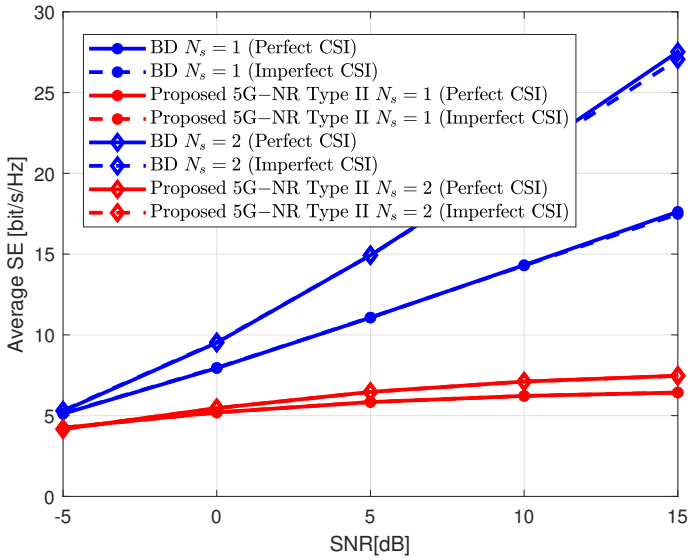

Figure 10. Average SE vs. SNR for MU-MIMO systems assuming $K = 2$, $N_t = 32, N_r = 4$, and different values of N_s. The gNB is implemented assuming a single-panel UPA with $N_1 = 4$, and $N_2 = 4$. Imperfect CSI is also assumed with $\sigma_e^2 = 10^{-3}$.

Figure 11 shows the sum-SE for different numbers of multiplexed UEs with the same PRBs. The higher the value of K, the higher the interference signals which limits the sum-SE performance increment. The correlation between the channel matrices of the different users influences the interference cancellation. Consequently, Type II codebook is only applicable to those users with low spatial correlation.

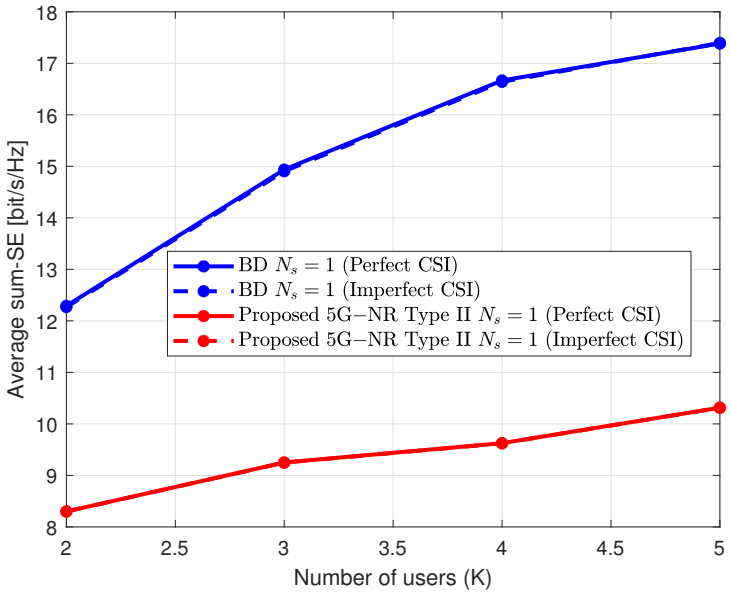

Figure 11. Average SE vs. number of users (K) for MU-MIMO systems assuming $N_t = 32, N_r = 4$, $N_s = 1$, and SNR=5dB. The gNB is implemented assuming a single-panel ULA with $N_1 = 16$, and $N_2 = 1$. Imperfect CSI is also assumed with $\sigma_e^2 = 10^{-3}$.

6. Conclusions

In this work, the performance bounds in terms of SE achieved by the 5G-NR precoding matrices are numerically evaluated for SU-MIMO and MU-MIMO systems. A detailed description of the 5G-NR precoding matrices implementation, covering both Type I and Type II codebooks, is provided according to TS 38-211 and TS 38-214. Numerical results are obtained for several antenna array configurations at the gNB. It can be noted that, although optimal SVD and BD provide a higher SE than the 5G-NR precoders, they require an unfeasible overhead that cannot be currently supported in real deployments. Indeed, the main goal of 5G-NR is to guarantee a tradeoff between performance, complexity, and feedback overhead. Furthermore, it is illustrated that the number of parallel layers that achieves the highest performance depends on the spatial properties of the channel matrix. In general, it was found that for low SNR values it is advisable to focus the beam on a single highly directional data stream than to multiplex several parallel layers. The effect of imperfect channel estimation is evaluated revealing that for low estimation error variance, in the order of 10^{-3}, the performance loss is negligible. In the case of MU-MIMO systems, the sum-SE is limited by the interference between near users. The use of advanced multi-antenna techniques to reach the demanding SE in 5G networks has been explored and validated, providing a comprehensive guide with useful insights for practical deployments. Future research can be conducted to explore the enhancements of mTRP and some resource allocation algorithms in conjunction with SU-MIMO and MU-MIMO. In addition, a further analysis of the upcoming releases of 3GPP regarding the CSI codebook may be considered.

Author Contributions: Conceptualization, D.A.U.V., H.O. and M.J.F.-G.G.; methodology, D.A.U.V., H.O. and M.J.F.-G.G.; software, D.A.U.V. and H.O.; validation, D.A.U.V., H.O. and M.J.F.-G.G.; formal analysis, D.A.U.V., H.O. and M.J.F.-G.G.; investigation, D.A.U.V., H.O. and M.J.F.-G.G.; resources, D.A.U.V., H.O. and M.J.F.-G.G.; data curation, D.A.U.V.; writing original draft preparation, D.A.U.V.; writing review and editing, D.A.U.V., H.O., M.J.F.-G.G. and A.F.; visualization, D.A.U.V., H.O. and

M.J.F.-G.G.; supervision, H.O., M.J.F.-G.G. and A.F.; project administration, M.J.F.-G.G. and A.F.; funding acquisition, M.J.F.-G.G. and A.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work has received funding from the European Union (EU) Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie ETN TeamUp5G, grant agreement No. 813391. Also, this work has been partially funded by the Spanish National project IRENE-EARTH (PID2020-115323RB-C33/AEI/10.13039/501100011033).

Data Availability Statement: The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest: The authors declare that there are no conflict of interest regarding the publication of this paper.

Abbreviations

The following Symbols are used in this manuscript:

$\mathbf{W}$	Precoding matrix
$\mathbf{W}_1$	Wideband precoding matrix
$\mathbf{w}_2$	Subband precoding vector
$\mathbf{U}$	Combining matrix
$\mathbf{H}$	Channel matrix
$\mathbf{s}$	Data symbols vector
$\mathbf{n}$	noise vector
R	Spectral-efficiency
K	Number of UEs
N_s	Number of multiplexed data streams
N_{SB}	Number of subbands
N_{sc}	Number of subcarriers
N_f	Number of subcarriers in each subband
N_{sc}^{RB}	Number of subcarriers in each physical resource block
N_{PRB}^{SB}	Number of physical resource blocks in each subband
N_t/N_r	Number of transmit/receive antennas
N_g	Number of antenna panels
N_1/N_2	Number of the dual-polarized antenna elements in the horizontal/vertical dimension
O_1/O_2	Oversampling factors in the horizontal/vertical dimension

References

1. Uwaechia, A.N.; Mahyuddin, N.M. A Comprehensive Survey on Millimeter Wave Communications for Fifth-Generation Wireless Networks: Feasibility and Challenges. *IEEE Access* **2020**, *8*, 62367–62414. [CrossRef]
2. Zaidi, A.; Athley, F.; Medbo, J.; Gustavsson, U.; Durisi, G.; Chen, X. *5G Physical Layer: Principles, Models and Technology Components*, 1st ed.; Academic Press: London, UK, 2018.
3. Ahmadi, S. *5G NR: Architecture, Technology, Implementation, and Operation of 3GPP New Radio Standards*; Academic Press: Cambridge, MA, USA, 2019.
4. Dahlman, E.; Parkvall, S.; Skold, J. *5G NR: The Next Generation Wireless Access Technology*, 2nd ed.; Academic Press: Cambridge, MA, USA, 2020.
5. Telatar, E. Capacity of Multi-antenna Gaussian Channels. *Eur. Trans. Telecommun.* **1999**, *10*, 585–595. . 460100604. [CrossRef]
6. Gesbert, D.; Kountouris, M.; Heath, R.W.; Chae, C.b.; Salzer, T. Shifting the MIMO Paradigm. *IEEE Signal Process. Mag.* **2007**, *24*, 36–46. [CrossRef]
7. Marzetta, T.L. Massive MIMO: An Introduction. *Bell Labs Tech. J.* **2015**, *20*, 11–22. [CrossRef]
8. Lin, Z.; Niu, H.; An, K.; Wang, Y.; Zheng, G.; Chatzinotas, S.; Hu, Y. Refracting RIS-Aided Hybrid Satellite-Terrestrial Relay Networks: Joint Beamforming Design and Optimization. *IEEE Trans. Aerosp. Electron. Syst.* **2022**, *58*, 3717–3724. [CrossRef]
9. Niu, H.; Lin, Z.; Chu, Z.; Zhu, Z.; Xiao, P.; Nguyen, H.X.; Lee, I.; Al-Dhahir, N. Joint Beamforming Design for Secure RIS-Assisted IoT Networks. *IEEE Internet Things J.* **2022**, 1.
10. Li, G.; Zeng, M.; Mishra, D.; Hao, L.; Ma, Z.; Dobre, O.A. Energy-Efficient Design for IRS-Empowered Uplink MIMO-NOMA Systems. *IEEE Trans. Veh. Technol.* **2022**, *71*, 9490–9500. [CrossRef]

11. Zeng, M.; Bedeer, E.; Dobre, O.A.; Fortier, P.; Pham, Q.V.; Hao, W. Energy-Efficient Resource Allocation for IRS-Assisted Multi-Antenna Uplink Systems. *IEEE Wirel. Commun. Lett.* **2021**, *10*, 1261–1265. [CrossRef]
12. Lin, Z.; Lin, M.; Wang, J.B.; de Cola, T.; Wang, J. Joint Beamforming and Power Allocation for Satellite-Terrestrial Integrated Networks With Non-Orthogonal Multiple Access. *IEEE J. Sel. Top. Signal Process.* **2019**, *13*, 657–670. [CrossRef]
13. Lin, Z.; An, K.; Niu, H.; Hu, Y.; Chatzinotas, S.; Zheng, G.; Wang, J. SLNR-based Secure Energy Efficient Beamforming in Multibeam Satellite Systems. *IEEE Trans. Aerosp. Electron. Syst.* **2022**, 1–4. [CrossRef]
14. Lee, H.H.; Ko, Y.C. Low Complexity Codebook-Based Beamforming for MIMO-OFDM Systems in Millimeter-Wave WPAN. *IEEE Trans. Wirel. Commun.* **2011**, *10*, 3607–3612. [CrossRef]
15. Wang, J.; Lan, Z.; Woo Pyo, C.; Baykas, T.; Sean Sum, C.; Rahman, M.; Gao, J.; Funada, R.; Kojima, F.; Harada, H.; et al. Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems. *IEEE J. Sel. Areas Commun.* **2009**, *27*, 1390–1399. [CrossRef]
16. Ren, Y.; Wang, Y.; Qi, C.; Liu, Y. Multiple-Beam Selection With Limited Feedback for Hybrid Beamforming in Massive MIMO Systems. *IEEE Access* **2017**, *5*, 13327–13335. [CrossRef]
17. Castellanos, M.R.; Raghavan, V.; Ryu, J.H.; Koymen, O.H.; Li, J.; Love, D.J.; Peleato, B. Channel-Reconstruction-Based Hybrid Precoding for Millimeter-Wave Multi-User MIMO Systems. *IEEE J. Sel. Top. Signal Process.* **2018**, *12*, 383–398. [CrossRef]
18. Heath, R.W.; González-Prelcic, N.; Rangan, S.; Roh, W.; Sayeed, A.M. An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems. *IEEE J. Sel. Top. Signal Process.* **2016**, *10*, 436–453. [CrossRef]
19. Wu, W.; Liu, D.; Hou, X.; Liu, M. Low-Complexity Beam Training for 5G Millimeter-Wave Massive MIMO Systems. *IEEE Trans. Veh. Technol.* **2020**, *69*, 361–376. [CrossRef]
20. Zhang, R.; Zhang, H.; Xu, W.; You, X. Subarray-Cooperation-Based Multi-Resolution Codebook and Beam Alignment Design for mmWave Backhaul Links. *IEEE Access* **2019**, *7*, 18319–18331. [CrossRef]
21. Albreem, M.A.; Habbash, A.H.A.; Abu-Hudrouss, A.M.; Ikki, S.S. Overview of Precoding Techniques for Massive MIMO. *IEEE Access* **2021**, *9*, 60764–60801. [CrossRef]
22. Schulz, B. *LTE Transmission Modes and Beamforming*; White Paper; Rohde & Schwarz: Munich, Germany, 2015.
23. 3GPP TS 36.213. LTE Evolved Universal Terrestrial Radio Access (E-UTRA), Physical Layer Procedures. Technical Specification. 2016. Available online: https://itecspec.com/archive/3gpp-specification-ts-36-213/ (accessed on 30 October 2022).
24. 3GPP TS 38.211. Physical Channels and Modulation. Technical Specification. 2022. Available online: https://itecspec.com/archive/3gpp-specification-ts-38-211/ (accessed on 30 October 2022).
25. 3GPP TS 38.214. Physical Layer Procedures for Data. Technical Specification. 2022. Available online: https://itecspec.com/archive/3gpp-specification-ts-38-214/ (accessed on 30 October 2022).
26. Hindy, A.; Mittal, U.; Brown, T. CSI Feedback Overhead Reduction for 5G Massive MIMO Systems. In Proceedings of the 2020 10th Annual Computing and Communication Workshop and Conference (CCWC), Vegas, NV, USA, 6–8 January 2020; pp. 0116–0120. [CrossRef]
27. Ahmed, R.; Tosato, F.; Maso, M. Overhead Reduction of NR type II CSI for NR Release 16. In Proceedings of the WSA 2019, 23rd International ITG Workshop on Smart Antennas, Vienna, Austria, 24–26 April 2019; pp. 1–5.
28. Suárez, L.; Ryabov, N.; Lyashev, V.; Sherstobitov, A. DFT Based Beam-Time Delay Sparse Channel Representation for Channel State Information (CSI) Compression in 5G FDD Massive MIMO Systems. In Proceedings of the 2018 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), Batumi, Georgia, 4–7 June 2018; pp. 1–5. [CrossRef]
29. Butovitsch, P.; Chapman, T.; Ghasemzadeh, F.; Hogan, B.; Karlsson, J.; Larsson, E.; Astely, D.; Göransson, B.; Friberg, C.; Jöngren, G.; et al. *Massive MIMO Handbook*; Ericsson AB. 2022. Available online: https://foryou.ericsson.com/Massive-MIMO-handbook-extended-version-download.html (accessed on 30 October 2022).
30. Enescu, M. *5G New Radio: A Beam-Based Air Interface*; Wiley & Sons, Limited, John: Hoboken, NJ, USA, 2020.
31. Asplund, H.; Astely, D.; Butovitsch, P.v. *Advanced Antenna Systems for 5G Network Deployments: Bridging the Gap Between Theory and Practice*; Academic Press: Cambridge, MA, USA, 2020.
32. Tang, H.; Yang, N.; Zhang, Z.; Du, Z.; Shen, J. *5G NR and Enhancements: From R15 to R16*; Elsevier: Amsterdam, Netherlands, 2021.
33. Ghosh, A.; Maeder, A.; Baker, M.; Chandramouli, D. 5G Evolution: A View on 5G Cellular Technology Beyond 3GPP Release 15. *IEEE Access* **2019**, *7*, 127639–127651. [CrossRef]
34. Riviello, D.G.; Di Stasio, F.; Tuninato, R. Performance Analysis of Multi-User MIMO Schemes under Realistic 3GPP 3-D Channel Model for 5G mmWave Cellular Networks. *Electronics* **2022**, *11*, 330. [CrossRef]
35. Henry, S.; Alsohaily, A.; Sousa, E.S. 5G is Real: Evaluating the Compliance of the 3GPP 5G New Radio System With the ITU IMT-2020 Requirements. *IEEE Access* **2020**, *8*, 42828–42840. [CrossRef]
36. Urquiza Villalonga, D.A.; OdetAlla, H.; Fernández-Getino García, M.J.; Flizikowski, A. Performance bounds with precoding matrices compliant with standardized 5G-NR for MIMO transmission. In Proceedings of the 2022 IEEE Conference on Standards for Communications and Networking (IEEE-CSCN), Rome, Italy, 20–23 June 2022.
37. 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]
38. 3GPP TR 38.901. Study on Channel Model for Frequencies from 0.5 to 100 GHz. Technical Specification. 2022. Available online: https://itecspec.com/archive/3gpp-specification-tr-38-901/ (accessed on 30 October 2022).

39. 3GPP TS 38.212. Multiplexing and Channel Coding. Technical Specification. 2022. Available online: https://itecspec.com/archive/3gpp-specification-ts-38-212/ (accessed on 30 October 2022).
40. Gomadam, K.; Cadambe, V.R.; Jafar, S.A. A Distributed Numerical Approach to Interference Alignment and Applications to Wireless Interference Networks. *IEEE Trans. Inf. Theory* **2011**, *57*, 3309–3322. [CrossRef]