

Search-Free Precoder Selection for 5G New Radio using Neural Networks

Talha Akyıldız and Tolga M. Duman
EEE Dept., Bilkent University, Ankara, Turkey
Email: {akyildiz, duman}@ee.bilkent.edu.tr

Abstract—We propose a search-free precoder selection method with neural networks motivated by the fact that large codebook sizes are adopted in 5G New Radio (5G-NR). The proposed method does not require an explicit codebook search unlike the traditional selection algorithms. Instead, it aims at finding the precoder matrix index that maximizes the corresponding channel capacity using a neural network directly. The network is trained off-line using extensive simulated data with the underlying channel statistics; however, the actual selection algorithm is based on simple calculations with the neural network, hence it is feasible for real time implementation. We demonstrate that the proposed search-free selection algorithm is highly efficient, i.e., it results in a performance very close to optimal precoder in the codebook while its complexity is significantly lower. Simulations with realistic channel models of 5G-NR corroborate these observations as well. We also show that pruning of the trained neural network gives a way to achieve further complexity reduction with a very small reduction in the system performance.

Index Terms— 5G New Radio, MIMO, channel state information, precoding, neural networks.

I. INTRODUCTION

LTE-A and 5G-NR employ a large number of antenna elements utilizing rectangular (2D) active antenna arrays at the base station (BS), which is also known as full-dimensional MIMO (FD-MIMO). Active antenna arrays provide capability of performing beamforming in both elevation and azimuth directions, and beamforming in both directions enables the BS to exploit the channel characteristics fully with the aid of channel state information (CSI) contributing to the system performance significantly [1], [2].

A comprehensive CSI framework is employed to support reliable and efficient communication in 5G-NR. CSI is reported by the user equipment (UE) to the BS with three important feedback parameters: 1) rank indicator (RI): the number of data layers that can be supported by the experienced channel; 2) precoder matrix index (PMI): the index of a precoder matrix chosen from a predefined codebook (for the given rank); 3) channel quality indicator (CQI): the index of the modulation scheme and code rate that result in a block error rate (BLER) lower than 0.1. The other feedback parameters include layer indicator (LI) and CSI-RS resource indicator (CRI) whose details can be found in [3], [4]. Selection of CSI parameters plays an important role in obtaining the best system performance for the specific channel conditions.

This research is supported by ASELSAN Inc. under the “5G Platform” project which is partly funded by the Scientific and Technological Research Council of Turkey (TUBITAK) under the project number 1160206.

In this paper, we focus on the PMI selection in 5G-NR which is performed via a codebook based approach whose specific implementation is determined by the UE. A common way is to calculate a metric for each candidate in the codebook and select the one with the best metric value, which requires an explicit search. The authors in [5] provide different precoder selection criteria, e.g., capacity, singular value and minimum mean square error, for choosing the precoder matrix from a given codebook. Mutual information (MI) based precoder selection algorithms over a codebook are proposed in [6], [7]. Precoder selection with low complexity distance metrics, i.e., phase and chordal distances, from a reference precoder computed by singular value decomposition (SVD) is presented in [8]. A post-equalization mutual information based precoder selection algorithm is given in [9].

Recently, there have also been different reduced complexity approaches for precoder selection. For example, a search-free algorithm is introduced in [10] exploiting the discrete-time Fourier transform (DFT) codebook structure, which is based on an iterative linear phase estimation (ILPE). In [11], a low complexity precoder search algorithm is presented for both low and high resolution CSI taking the advantage of dual-stage codebook structure in 5G-NR. The work in [12] proposes computationally efficient precoder matrix selection adapting longer intervals for PMI feedback according to a calculated metric, such as the mutual information. In [13], the authors propose a neural network based approach to obtain the optimal channel sensing vectors and to construct the RF beamforming/combining vectors using the sensed channel for millimeter wave (mmWave) massive MIMO systems.

In 5G-NR, the precoder codebook sizes may be very large due to the high number of transmit antenna ports. For example, for the Type I single-panel codebook, the number of PMI candidates may be up to 2048. This makes an explicit codebook search, i.e., calculating a suitable metric for each candidate and selecting the optimal one, infeasible due to the increased complexity. Motivated by this, in this paper, we propose a search-free precoder selection algorithm that employs neural networks. We develop a simple but effective network architecture and show that the selection of a precoder matrix index is indeed feasible with this approach. Specifically, we train the network to find the channel capacity maximizing PMI from the codebook, and show that the proposed algorithm provides comparable results to the search-based algorithms; or, it may even outperform some of them, while offering

significantly lower complexity.

The remainder of the paper is organized as follows. In Section II, we describe the system model and specify the codebook structure adopted in 5G-NR. Section III demonstrates the proposed neural network based precoder selection method. Section IV illustrates our numerical results along with a detailed comparison with other approaches. Finally, Section V concludes the paper.

Notation: Throughout this paper, lowercase letters such as x , lowercase boldface letters such as \mathbf{x} and uppercase boldface letters such as \mathbf{X} denote scalars, vectors and matrices, respectively. The conjugate transpose (Hermitian) of \mathbf{X} is denoted by \mathbf{X}^H , and $\det(\cdot)$ denotes the determinant operator. The identity matrix of size $M \times M$ is represented by \mathbf{I}_M , and the zero matrix of size $M \times N$ is denoted by $\mathbf{0}_{M \times N}$. The circularly symmetric complex Gaussian random vector with mean vector $\boldsymbol{\mu}$ and covariance matrix \mathbf{C} is denoted by $\mathcal{CN}(\boldsymbol{\mu}, \mathbf{C})$. The Kronecker and Hadamard products are represented by symbols \otimes and \odot , respectively. Lastly, the real and imaginary parts of matrix \mathbf{X} are denoted by $\text{Re}(\mathbf{X})$ and $\text{Im}(\mathbf{X})$, respectively.

II. SYSTEM MODEL

We consider a single-user MIMO system where a BS transmits independently modulated N_l layers of data symbols using the same constellation to a UE. The symbols are represented by $\mathbf{s} = [s_1, s_2, \dots, s_{N_l}]^T$. The symbol vector \mathbf{s} is mapped from data layers to antenna ports by a precoder matrix chosen from a predefined codebook. The precoder matrix is denoted by $\mathbf{W} \in \mathbb{C}^{N_t \times N_l}$, where N_t is the number of antenna ports, with $N_t \geq N_l$. After the precoding operation, the mapped symbols are transmitted through the antenna ports to a receiver equipped with N_r antennas. Assuming perfect timing and synchronization, the received signal vector can be written as

$$\mathbf{y} = \mathbf{H}\mathbf{W}\mathbf{s} + \mathbf{n}. \quad (1)$$

where $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ denotes the channel matrix, $\mathbf{y} \in \mathbb{C}^{N_r \times 1}$ and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}_{N_r \times 1}, \sigma_n^2 \mathbf{I}_{N_r})$ denotes the additive white Gaussian noise (AWGN) vector.

In 5G-NR, a precoder matrix is selected from a codebook denoted by \mathcal{W} as specified in [4]. The codebook can be classified into two types: Type I and II. Type I codebook is designed for single-user MIMO with low resolution, on the other hand, Type II codebook targets multi-user MIMO providing higher resolution to suppress the inter-user interference. In this paper, our main focus is on the single-user MIMO scenario, and hence, only the Type I codebook is considered. Type I codebook structure supports single and multi antenna panels where each panel consists of N_1 horizontal antennas and N_2 vertical antennas with cross-polarization. Hence, the total number of transmit antenna ports for each panel is $N_t = 2N_1N_2$. Type I codebooks are generated by the Kronecker products of DFT vectors which are defined as

$$\mathbf{h}_1 = \begin{bmatrix} 1 & e^{j\frac{2\pi l}{O_1 N_1}} & \dots & e^{j\frac{2\pi l(N_1-1)}{O_1 N_1}} \end{bmatrix}^T,$$

$$\begin{aligned} \tilde{\mathbf{h}}_1 &= \begin{bmatrix} 1 & e^{j\frac{4\pi l}{O_1 N_1}} & \dots & e^{j\frac{4\pi l(N_1/2-1)}{O_1 N_1}} \end{bmatrix}^T, \\ \mathbf{v}_m &= \begin{bmatrix} 1 & e^{j\frac{2\pi m}{O_2 N_2}} & \dots & e^{j\frac{2\pi m(N_2-1)}{O_2 N_2}} \end{bmatrix}^T, \end{aligned} \quad (2)$$

where \mathbf{h}_1 and $\tilde{\mathbf{h}}_1$ represent the beam in the horizontal direction; while \mathbf{h}_1 has a higher granularity and narrow beams compared to $\tilde{\mathbf{h}}_1$. The vector \mathbf{v}_m represents the beam in the vertical direction. The parameters O_1 and O_2 are the over-sampling factors for horizontal and vertical directions, respectively. In addition to the DFT vectors, for antenna polarization and co-phasing, the variables $\phi_n = e^{j\pi n/2}$ and $\theta_p = e^{j\pi p/4}$ are specified as well. Using the defined DFT vectors, the Kronecker product vectors can be written as $\mathbf{b}_{1,m} = \mathbf{h}_1 \otimes \mathbf{v}_m$ and $\tilde{\mathbf{b}}_{1,m} = \tilde{\mathbf{h}}_1 \otimes \mathbf{v}_m$.

The Type I single-panel codebook can support 8 layers of transmission while this number decreases to 4 when multiple panels are adopted. The codebook generation is specific to the number of layers, and it is described in detail in [4]. We only illustrate an example for the scenario when there is only a single layer, $N_l = 1$, noting that the codebook generation is analogous for $N_l > 1$. We also remark that for layer 1 and 2, the precoder codebook contains two different modes. Codebook mode 1 allows for higher granularity in horizontal and vertical directions for wide-band, whereas codebook mode 2 has higher resolution for sub-band. The codebook structure for 1 layer transmission considering mode 1 is given in Table I, where the indexes $i_{1,1}$, $i_{1,2}$ and i_2 are employed for its generation. It can be seen that the parameters l , m and n are determined from $i_{1,1}$, $i_{1,2}$ and i_2 , respectively.

TABLE I: Codebook generation in 5G-NR Type-I single-panel for 1 layer.

Codebook mode 1			
$i_{1,1}$	$i_{1,2}$	i_2	
$0, 1, \dots, N_1 O_1 - 1$	$0, 1, \dots, N_2 O_2 - 1$	$0, 1, 2, 3$	$\mathbf{W}_{i_{1,1}, i_{1,2}, i_2}$
where $\mathbf{W}_{1,m,n} = \sqrt{\frac{1}{N_t}}$			$\begin{bmatrix} \mathbf{b}_{1,m} \\ \phi_n \odot \mathbf{b}_{1,m} \end{bmatrix}$

5G-NR utilizes a dual-stage codebook structure where the precoding matrix can be expressed as a product of two matrices, i.e., $\mathbf{W} = \mathbf{W}_1 \mathbf{W}_2$. The first matrix \mathbf{W}_1 represents the long-term channel characteristics (wide-band), while \mathbf{W}_2 captures the short-term channel characteristics (sub-band). \mathbf{W}_1 contains multiple beam directions, and \mathbf{W}_2 matrix picks the subset of beam directions and also performs phase shifting. The selection of beam directions is only available for codebook mode 2 while phase shifting is common for both modes (1-2). \mathbf{W}_1 and \mathbf{W}_2 matrices are dependent on the indexes $i_{1,k}$ for $k \in \{1, 2, 3\}$ and i_2 , respectively. For layer 1, \mathbf{W}_1 and \mathbf{W}_2 can be written as

$$\mathbf{W}_1 = \begin{bmatrix} \mathbf{b}_{1,m} & \mathbf{0} \\ \mathbf{0} & \mathbf{b}_{1,m} \end{bmatrix}, \quad \mathbf{W}_2 = \begin{bmatrix} 1 \\ \phi_n \end{bmatrix}.$$

We also present the supported configurations of (N_1, N_2) and (O_1, O_2) in 5G-NR with the corresponding codebook

sizes, $|\mathcal{W}|$, for the number of layers up to 4 considering Type I single-panel (mode 1) in Table II. It can be seen that the codebook size depends on the number of transmitter antenna ports as well as their placements, and they are specific to the number of layers. We highlight that large codebook sizes necessitate a low complexity precoder selection approach.

TABLE II: Codebook sizes adopted in 5G-NR with the supported configurations of (N_1, N_2) and (O_1, O_2) for different number of layers.

(N_1, N_2)	(O_1, O_2)	$ \mathcal{W} $ ($N_l = 1$)	$ \mathcal{W} $ ($N_l = 2$)	$ \mathcal{W} $ ($N_l = 3, 4$)
(2,1)	(4,1)	32	32	16
(2,2)	(4,4)	256	512	384
(4,1)	(4,1)	64	128	96
(3,2)	(4,4)	384	768	768
(6,1)	(4,1)	96	192	192
(4,2)	(4,4)	512	1024	512
(8,1)	(4,1)	128	256	128
(4,3)	(4,4)	768	1536	768
(6,2)	(4,4)	768	1536	768
(12,1)	(4,1)	192	384	192
(4,4)	(4,4)	1024	2048	1024
(8,2)	(4,4)	1024	2048	1024
(16,1)	(4,1)	256	512	256

III. PROPOSED NEURAL NETWORK BASED PRECODER SELECTION

In this section, we propose a precoder selection algorithm that uses learning tools, specifically, a feed-forward neural network with a single hidden layer. We model the selection of the precoder matrix index which maximizes the channel capacity as a classification problem. Particularly, we predict the precoder matrix index for a given channel realization \mathbf{H} .

In the following, we first develop a simple yet effective neural network framework and explain the proposed architecture. Then, we describe the training procedure using supervised learning and provide the training results. Lastly, we compare the complexity of the proposed structure with those of the existing precoder selection algorithms.

A. Proposed Neural Network Architecture

The proposed neural network architecture is illustrated in Fig. 1. At the preprocessing stage, before feeding the channel

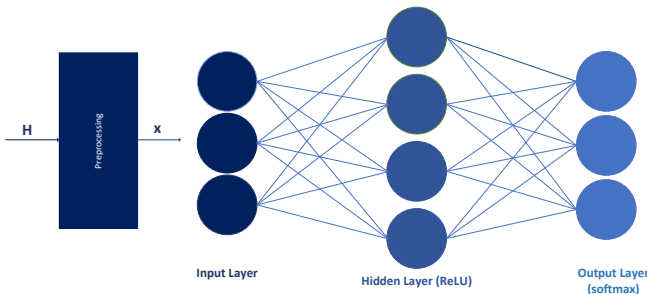


Fig. 1: The proposed neural network architecture.

matrix into the neural network, we first convert it into a vector denoted by \mathbf{x} . Specifically, we vectorize \mathbf{H} concatenating its real and imaginary parts and obtain $\mathbf{x} \in \mathbb{R}^{2N_t N_r \times 1}$. The number of input units, N_i , is equal to the length of \mathbf{x} . We denote the number of units in the hidden and output layers by N_h and N_o , respectively. It should be noted that the number of output units is determined by the codebook size, and hence, $N_o = |\mathcal{W}|$. Let \mathbf{W}_h , \mathbf{W}_o and \mathbf{b}_h , \mathbf{b}_o represent the weight matrices and bias vectors for the hidden and output layers, respectively.

We use rectified linear unit (ReLU) as the activation function at the hidden layer with the aim of enabling the neural network to approximate the relevant non-linear function, i.e., the channel capacity, sufficiently well. Moreover, softmax is utilized at the output layer which is commonly used for multi-class classification. The softmax operation converts the input values into a probability distribution over possible output classes. The relationship among the input, hidden and output layers can be written as

$$\begin{aligned} \mathbf{h} &= \text{ReLU}(\mathbf{W}_h \mathbf{x} + \mathbf{b}_h), \\ \mathbf{o} &= \text{softmax}(\mathbf{W}_o \mathbf{h} + \mathbf{b}_o), \end{aligned} \quad (3)$$

where $\mathbf{W}_h \in \mathbb{R}^{N_h \times N_i}$, $\mathbf{W}_o \in \mathbb{R}^{N_o \times N_h}$, $\mathbf{b}_h \in \mathbb{R}^{N_h \times 1}$ and $\mathbf{b}_o \in \mathbb{R}^{N_o \times 1}$. The activation function ReLU and softmax operation are element-wise functions, i.e., they perform the same operation for each element of the input vector (denoted by a_i), and they are defined as

$$\text{ReLU}(a_i) = \max(0, a_i), \quad \text{softmax}(a_i) = \frac{e^{a_i}}{\sum_{j=1}^{N_o} e^{a_j}}, \quad (4)$$

where the calculations are performed over all possible i values. As a last step, the most likely precoder index, i.e., the output of the neural network, is the one with the maximum probability produced by the softmax operation, i.e.,

$$\hat{i} = \underset{i \in \{1, 2, \dots, N_o\}}{\text{argmax}} \ o_i, \quad (5)$$

where o_i is the i -th element of \mathbf{o} .

B. Training Procedure

We now explain the training procedure of the proposed neural network described in the previous subsection. The objective of the training is to learn the weight and bias parameters which minimize the given loss function. By doing that, our aim is to predict the true precoder index for any given channel matrix using supervised learning. In the training procedure, we use channel matrices whose elements are independent and identically distributed (i.i.d.) with zero mean and unit variance complex Gaussian random variables, i.e., $\mathcal{CN}(0, 1)$ ¹. The true index for each channel matrix, i.e., the precoder matrix index such that the channel capacity is maximized, is selected as

$$i^* = \underset{i \in \{1, 2, \dots, |\mathcal{W}|\}}{\text{argmax}} \ \log_2 \det \left(\mathbf{I}_{N_t} + \frac{1}{\sigma_n^2} (\mathbf{W}^i)^H \mathbf{H}^H \mathbf{H} \mathbf{W}^i \right). \quad (6)$$

¹We note that while the training is performed with this specific model, the trained network can be used with other channel models as well.

where \mathbf{W}^i denotes the i -th precoder matrix in the codebook \mathcal{W} . The true label vector, $\mathbf{d} \in \mathbb{R}^{|\mathcal{W}| \times 1}$, is formed in such a way that it has 1 entry corresponding to the index of the true label, i^* , and zero entries for all the other indexes. We employ the cross entropy loss function which can be written in terms of \mathbf{d} and \mathbf{o} as

$$J = - \sum_{m=1}^{|\mathcal{W}|} d_m \log(o_m), \quad (7)$$

where d_m and o_m denote the m -th element of \mathbf{d} and \mathbf{o} , respectively.

The loss function is minimized using a specified optimization algorithm with respect to the weight and bias terms, i.e., \mathbf{W}_h , \mathbf{W}_o , \mathbf{b}_h and \mathbf{b}_o , using the gradients calculated through a back-propagation algorithm.

C. Training Results

We use 10^7 channel realizations, and divide them into two parts: 8×10^6 for training, and 2×10^6 for validation. As a specific example, we choose $(N_1, N_2) = (4, 4)$ and $(O_1, O_2) = (4, 4)$ and set the number of layers to 1. The number of candidate precoder matrices for this specific configuration is 1024 (see Table II). In this configuration, the number transmit antenna ports corresponds to 32 using the relationship $N_t = 2N_1N_2$, and we select the number of receive antenna ports as 1. We determine the number of hidden units as 1024 since less number of units results in a performance loss and higher number of units do not necessarily improve it. The number of output units is equivalent to the codebook size $|\mathcal{W}|$.

We use a stochastic gradient descent with momentum (SGDM) optimizer with an initial learning rate $\eta = 0.1$ and a momentum factor $\gamma = 0.9$ by noting that we do not observe significant performance gain with the other optimizers, e.g., RmsProp [14] and Adam [15]. We choose the mini batch size as 8192 and employ the validation check to avoid over-fitting. The corresponding accuracy and cross entropy loss values for the training and validation sets over the iterations (epochs) are depicted in Fig. 2. We observe that the loss is decreasing through the iterations, and after a certain number of them, it is nearly a constant. We also observe that the accuracy increases as the network updates the weights and bias terms with SGDM, and the validation accuracy is around 73% after all the iterations are complete. This result is quite fulfilling considering that there are 1024 classes, and many of misclassifications result in precoder indexes whose corresponding capacity is quite close to the true one, i.e., the capacity loss is almost negligible.

D. Complexity Analysis

We compare the complexity of the proposed method with the search based algorithms employing capacity, chordal distance and phase distance [8]. We first note that, in the phase distance and chordal distance based selection algorithms, the reference precoder needs to be computed via SVD of the channel matrix $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$, which requires $4N_t^2N_l + 22N_l^3$ operations which is computationally expensive [16]. Then,

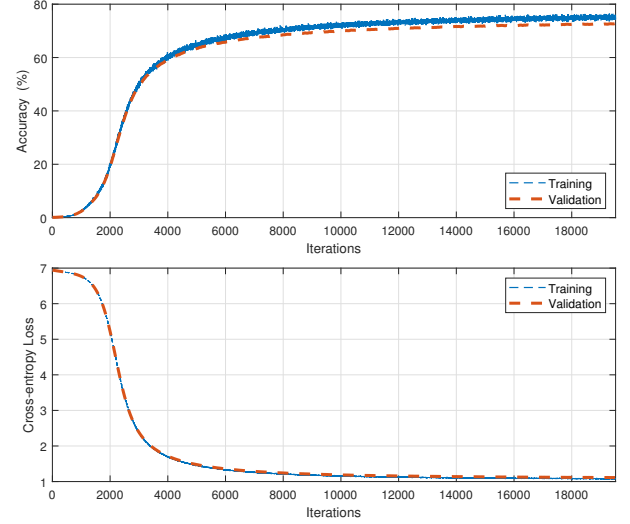


Fig. 2: The accuracy and cross-entropy loss values for training and validation data sets over iterations.

the corresponding distance metrics are calculated per each candidate in the codebook for a given reference precoder. On the other hand, the capacity based selection algorithm does not require an SVD operation, while the channel capacity needs to be computed for each precoder matrix using (6).

The required operations in terms multiplications and additions for each algorithm are given in Table III ². In our proposed method, the neural network involves two multiplications between the input and output layers, with two weight matrices, \mathbf{W}_h and \mathbf{W}_o . The summation terms stem from the bias terms and the softmax operation. We note that the required operations in Table III are per each candidate in the codebook for the other algorithms while this is not the case for the proposed method. Moreover, our algorithm only contains operations with real numbers whereas the other methods operate with complex numbers which increases their complexity further.

Note that we can also prune the neural network to further reduce the complexity while keeping the performance close to the non-pruned one. In a pruned neural network, the elements in the weight matrices and bias vectors with magnitude lower than a certain threshold are set to 0. That is, the weight matrices and bias vectors become sparse and the required number of operations are greatly reduced. In our numerical experiments, we have observed that a 90% sparsification with pruning results in almost no performance loss.

A direct complexity comparison of the algorithms is not

TABLE III: Complexity comparison of different selection algorithms.

Algorithm	Multiplication	Summation/Subtraction
Phase Distance	-	$2(N_t - 1)N_l + N_l^2N_t + N_lN_l!$
Chordal Distance	$N_l^2N_t$	N_l^2
Capacity	$N_rN_tN_l + N_l^2N_r + N_l^3$	N_l
Neural Network	$2N_rN_tN_h + N_hN_o$	$N_h + 2N_o$

²For the details of the phase and chordal distance computations, see [8].

feasible due to the different type of operations and parameters being used. Thus, we also provide computation times of the algorithms for an exemplary codebook of size 1024. We adopt the parameters $N_t = 1$, $N_r = 32$ and $N_s = 1$, and choose the number of units for the hidden and output layers of the neural network as $N_h = N_o = 1024$. The corresponding Matlab implementation results are given in Table IV. We observe that the capacity based selection has the highest computation time due to the required number of complex multiplications. On the other hand, the phase distance and chordal distance based selections benefit from the low complexity distance metrics compared to the channel capacity, and offer reduced complexity despite of the required SVD operation. Most importantly, we observe that the newly proposed neural network based solution is more efficient than the existing search-based approaches, and with pruning, it is possible to further decrease its computation time. We also note that the benefits of the neural network based approach will be even more pronounced for larger codebook sizes and potentially larger number of antennas and layers.

IV. NUMERICAL RESULTS

In this section, we illustrate the performance of the proposed precoder selection algorithm in comparison with the existing ones via simulations. We first consider the basic simulation setup, i.e., Rayleigh block fading channel model, and then extend our examples to the realistic 5G channel models.

In the first simulation setup, we use QPSK modulation scheme, and take the codebook parameters as $(N_1, N_2) = (4, 4)$ and $(O_1, O_2) = (4, 4)$ for Type I single-panel precoding. The number of data layers is chosen as 1, the number of transmitter and receiver antenna ports are 32 and 1, respectively. The linear minimum mean-square error (MMSE) decoder is utilized at the receiver. We assume Rayleigh block fading channel model, where the channel is constant during each sub-frame and independent of each other. We assume that the channel estimation is ideal and the PMI feedback is sent to the transmitter with zero delay. In other words, the precoder matrix index is chosen before the transmission starts.

Fig. 3 illustrates the capacity and bit error rate performance of different PMI selection algorithms over various SNR values. For each SNR value, 10^3 channel realizations are simulated. For comparison purposes, we consider capacity, chordal distance and phase distance based selection algorithms which

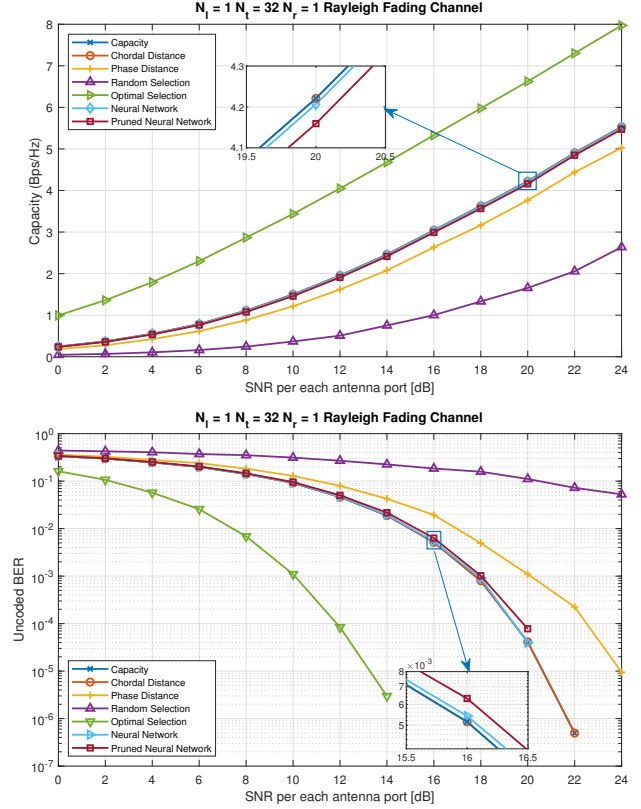


Fig. 3: Capacity and bit error performance of different precoder selection algorithms with Rayleigh fading channel model.

require exhaustive search over the precoder codebook. Note that an alternative search-free precoder selection algorithm is proposed in [10]. While this algorithm is of low complexity, its performance is not comparable with the approaches that are based on a search over the codebook using different criteria such as capacity maximization. Due to their superior performance, we only compare our proposed approach with the search-based solutions. We also include the optimal selection algorithm that obtains the precoder matrix via the SVD of the channel matrix and a random selection algorithm which chooses the precoder matrix index randomly without calculating any metric.

We observe from Fig. 3 that the newly proposed algorithm provides almost identical performance (even with pruning) compared to the capacity and chordal distance based approaches. Indeed, the capacity loss is almost negligible while providing better results than the phase distance based selection. Obviously, the optimal selection is superior in terms of both metrics taking advantage of precoders that are not available in the codebook.

In the second simulation setup, we adopt the same parameters, but change the channel model to a realistic 5G model with delay profile TDL-A [17]. We assume that there is no correlation among the transmitter antennas and the delay spread is 300 ns. The maximum Doppler shift is 10 Hz and the sub-carrier spacing is set to 15 kHz. The total bandwidth is equal to 5 MHz and the number of available resource blocks

TABLE IV: Computation time comparison of different selection algorithms for a specific example.

Algorithm	Code Run Time
Phase Distance	≈ 1.5 ms
Chordal Distance	≈ 2 ms
Capacity	≈ 3.5 ms
Neural Network	≈ 1.1 ms
Pruned Neural Network	≈ 0.6 ms

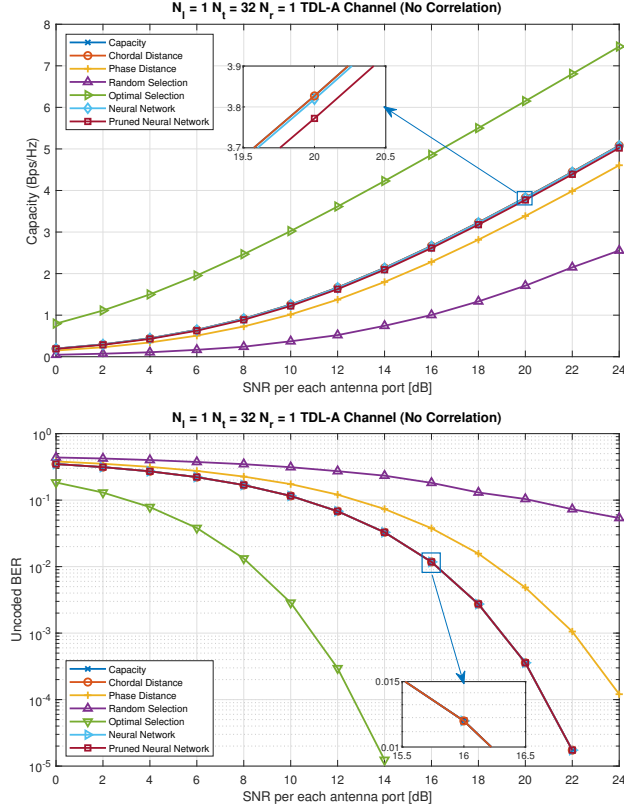


Fig. 4: Capacity and bit error performance of different precoder selection algorithms with 5G channel model (TDL-A).

(RB) is 20. 12 sub-carriers are allocated for each RB on frequency and 14 symbols are allocated across time. Therefore, the corresponding number of resource elements is 3360 per sub-frame. The precoder matrix is selected by averaging the channel matrix over all resource elements for each algorithm, i.e., a single precoding matrix is selected for the averaged channel matrix, and it is applied for all the resource elements.

Fig. 4 depicts the capacity and BER values of the different selection algorithms. The observations are similar to the previous example: the proposed algorithm achieves almost same performance with the capacity and chordal distance based selections for this realistic 5G channel model as well. This shows that even though the network is trained with a different channel model (i.e., i.i.d. Rayleigh fading), it is capable of selecting favorable precoder matrix indexes for the realistic channel models as well. We also observe that the gap between the capacity based selection and the proposed model is less than the previous one for both capacity and BER. We emphasize that, as detailed in the previous section, the neural network based solution obtains these very competitive performance results with significantly lower complexity than the existing alternatives.

V. CONCLUSIONS

We have considered the problem of selecting the precoder matrix from a given codebook, specifically, for 5G-NR. Different from the existing PMI selection methods, we propose

a search-free selection algorithm with a neural network. We show that with a suitable training procedure and sufficient training time, suitably designed and trained neural networks are capable of predicting the precoder matrix index for a chosen metric with excellent performance even when the number of candidates is large. We demonstrate that the proposed solution is computationally more efficient compared to the existing (search-based) algorithms, and it is possible to further reduce its complexity by pruning the network with almost no performance loss. We also perform simulations to demonstrate the effectiveness of the proposed method and verify that the proposed search-free approach can achieve very close performance to the search-based solutions with lower complexity.

REFERENCES

- [1] E. Onggosanusi *et al.*, "Modular and high-resolution channel state information and beam management for 5G New Radio," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 48–55, Mar. 2018.
- [2] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.
- [3] 3GPP, "NR; Multiplexing and channel coding," Tech. Rep. 38.212, Jan. 2019, v15.4.0. [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/38_series/38.212/
- [4] 3GPP, "NR; Physical layer procedures for data," Tech. Rep. 38.214, Dec. 2018, v15.4.0. [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/38_series/38.214/
- [5] D. J. Love and R. W. Heath, "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2967–2976, Aug. 2005.
- [6] T. Tang, R. Doostnejad, and T. J. Lim, "Mean mutual information per coded bit based precoding in MIMO-OFDM systems," in *Proc. IEEE Veh. Technol. Conf.*, Ottawa, Canada, Sep. 2010.
- [7] S. Schwarz, M. Wrulich, and M. Rupp, "Mutual information based calculation of the precoding matrix indicator for 3GPP UMTS/LTE," in *Proc. Int. ITG Workshop on Smart Antennas*, Bremen, Germany, Feb. 2010.
- [8] F. Penna, H. Cheng, and J. Lee, "Low complexity precoder selection for FD-MIMO systems," in *Proc. IEEE Veh. Technol. Conf.*, Montreal, Canada, Sep. 2016.
- [9] S. Schwarz, C. Mehlführer, and M. Rupp, "Calculation of the spatial preprocessing and link adaption feedback for 3GPP UMTS/LTE," in *Proc. Wirel. Adv.*, London, UK, Jun. 2010.
- [10] F. Penna, H. Cheng, and J. Lee, "A search-free algorithm for precoder selection in FD-MIMO systems with DFT-based codebooks," in *Proc. IEEE Veh. Technol. Conf.*, Toronto, Canada, Sep. 2017.
- [11] B. Mondal, V. Sergeev, A. Sengupta, and A. Davydov, "5G-NR (New Radio) CSI computation algorithm and performance," in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, USA, Oct. 2018.
- [12] M. Mussbah, S. Pratschner, S. Schwarz, and M. Rupp, "Computationally efficient limited feedback for codebook-based FD-MIMO precoding," in *Proc. Int. ITG Workshop on Smart Antennas*, Vienna, Austria, Apr. 2019.
- [13] X. Li and A. Alkhateeb, "Deep learning for direct hybrid precoding in millimeter wave massive MIMO systems," *arXiv:1905.13212*, 2019. [Online]. Available: <http://arxiv.org/abs/1905.13212>
- [14] T. Tieleman and G. Hinton, "Lecture 6.5—RmsProp: Divide the gradient by a running average of its recent magnitude," COURSE: Neural Networks for Machine Learning, 2012.
- [15] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Rep. (ICLR)*, San Diego, USA, May 2015.
- [16] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. The Johns Hopkins University Press, 1996.
- [17] 3GPP, "NR; Study on channel model for frequencies from 0.5 to 100 GHz," Tech. Rep. 38.901, Sep. 2019, v15.1.0. [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/38_series/38.901/