Learning When to Learn: Distortion-Aware GNN Precoding for Multi-Carrier Large Antenna Systems with Adaptive Precoder Selection

Thomas Feys, Liesbet Van der Perre, Gilles Callebaut, François Rottenberg
KU Leuven, Dept. ESAT-WaveCore, Dramco, B-9000 Ghent, Belgium

***Abstract*—In large-antenna systems, operating power amplifiers (PAs) close to saturation is key to achieving high energy efficiency. However, this operating regime inevitably introduces nonlinearities, whose resulting distortion can coherently combine at the user location and degrade system performance. Distortion-aware precoding techniques have been proposed to send this distortion in non-user directions. These distortion-aware precoders are typically designed under the assumption of a frequency flat channel. In contrast, we assume a multi-carrier, single-user downlink system that experiences frequency-selectivity. In multi-carrier systems, the non-linearity of the PA produces inter-modulation products, creating inter-carrier interference so that the precoding problem becomes coupled across subcarriers. To tackle this, we propose a graph neural network (GNN)-based joint precoding method that optimizes precoding vectors across all subcarriers while accounting for distortion. Simulation results indicate that this joint GNN precoder achieves stable rates across a wide range of channel conditions, outperforming classical maximum ratio transmission and per-subcarrier GNN approaches, albeit at increased computational complexity. Additionally, we show that the need for such a method depends strongly on the channel conditions. As such, we propose a convolutional neural network-based classifier that selects the appropriate precoding method based on the current channel conditions, while balancing the achievable rate and complexity.**

***Index Terms*—Power Amplifier, OFDM, Non-Linear Distortion, Precoding, Precoding Selection, Neural Networks**

I. Introduction

The power amplifier (PA) is one of the major energy consumers within a base station (BS) [1]. Consequently, its energy efficiency should be maximized. Unfortunately, there is a trade-off to be made. Operating the PA closer to saturation improves its energy efficiency. However, this introduces non-linear distortion, limiting the communication performance. As such, many works investigate how to overcome this non-linear distortion to improve the energy efficiency of the PAs [2]–[6]. An interesting avenue is to incorporate knowledge of the distortion into the precoder design [4]–[6]. This allows for the spatial suppression of the distortion in the user directions. These precoders are typically designed under frequency-flat/single-carrier channel conditions. Similarly, classical precoder design is done for a single flat carrier. Extending these precoders to orthogonal frequency division multiplexing (OFDM) systems is trivial, as they are applied independently per subcarrier. Unfortunately, independent per-subcarrier processing is not optimal when a non-linear PA is present. The non-linearity of the PA produces inter-modulation products, creating inter-carrier interference so that the precoding problem becomes coupled across subcarriers. This typically results in a degradation of the performance of these distortion-aware precoders. Neural network (NN)-based precoders have been developed for multi-carrier systems [7]. However, conventionally, no spectral leakage occurs as only linear components are considered. In [8], the performance of a massive MIMO system is evaluated in the multi-carrier case, where non-linear digital-to-analog converters are present, causing spectral leakage. However, no distortion-aware precoding techniques were proposed. In this work, we extend distortion-aware precoders to the multi-carrier case by introducing a graph neural network (GNN) which jointly computes the precoding vectors for all subcarriers while taking distortion into account. It is shown that this joint precoding delivers stable performance when the channel conditions vary, as compared to distortion-aware precoders developed for narrowband channels. Additionally, it is shown that the gain with respect to the classical maximum ratio transmission (MRT) precoder varies depending on the frequency-selectivity of the channel. This improved performance, however, comes at the cost of increased computational complexity. As such, we propose the use of a classifier that selects the appropriate precoding scheme depending on the current channel conditions. This classifier takes into account the achievable rate/complexity trade-off.

Partially funded by 6GTandem part of SNS JU under the European Union's Horizon Europe research and innovation program, Grant 101096302.

Co-funded by the European Union under Grant 101191936. Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of all 6GTandem/SUSTAIN-6G consortium parties nor those of the European Union or the SNS-JU (granting authority). Neither the European Union nor the granting authority can be held responsible for them. All code is available in this GitHub repository.

Notations: Vectors and matrices are denoted by bold lowercase and uppercase letters ($\mathbf{a}$, $\mathbf{A}$). Conjugate, transpose, and Hermitian transpose are $(\cdot)^*$, $(\cdot)^\intercal$, and $(\cdot)^H$; expectation is $\mathbb{E}[\cdot]$. The $M \times M$ identity is $\mathbf{I}_M$, diagonal of $\mathbf{A}$ is $\mathrm{diag}\,(\mathbf{A})$, and trace is $\mathrm{tr}\,(\mathbf{A})$. Matrix entries are $[\mathbf{A}]_{i,j}$ and vector entries $[\mathbf{a}]_i$ or a_i. A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ has nodes $\mathcal{V}$ and edges $\mathcal{E}$; edge $(a, b) \in \mathcal{E}$ joins nodes a and b, and the neighborhood of a is $\mathcal{N}(a) = \{b \in \mathcal{V} : (a, b) \in \mathcal{E}\}$.

II. Problem Formulation

A. System Model

We consider a downlink system with M antennas at the BS, one single antenna user, and Q subcarriers. The precoded symbols at subcarrier q are denoted as $\mathbf{x}_q = \mathbf{w}_q s_q$, where $\mathbf{x}_q \in \mathbb{C}^{M\times 1}$, $\mathbf{w}_q \in \mathbb{C}^{M\times 1}$ is the precoding vector at carrier q and $s_q \in \mathbb{C}$ is the user symbol at carrier q. Before applying the PA nonlinearity, the inverse discrete Fourier transform (IDFT) is applied to obtain the time domain signal

$$\tilde{x}_m[n] = \mathcal{F}^{-1}(x_{m,q}) = \frac{1}{\sqrt{Q}} \sum_{q=0}^{Q-1} x_{m,q} e^{\jmath 2\pi \frac{qn}{Q}} \tag{1}$$

where $\mathcal{F}^{-1}(\cdot)$ represents the IDFT and the time-domain signal is indicated as $\tilde{x}_m[n]$ with n being the discrete time-domain index. Afterward, the time domain signal is amplified as $\tilde{y}_m[n] = \phi_m(\tilde{x}_m[n])$, where $\phi_m(\cdot)$ denotes the nonlinear transformation caused by the PA at antenna m. The nonlinear amplifier is modeled as a complex-valued polynomial of order $2D+1$, which captures both amplitude modulation to amplitude modulation (AM/AM) and amplitude modulation to phase modulation (AM/PM) distortion [9]

$$\tilde{y}_m[n] = \phi_m(\tilde{x}_m[n]) = \sum_{d=0}^{D} a_{2d+1} \tilde{x}_m[n] |\tilde{x}_m[n]|^{2d}. \tag{2}$$

The amplified signal is transformed via the discrete Fourier transform (DFT) to analyze it in the frequency domain

$$y_{m,q} = \mathcal{F}(\tilde{y}_m[n]) = \frac{1}{\sqrt{Q}} \sum_{n=0}^{Q-1} \tilde{y}_m[n] e^{-\jmath 2\pi \frac{qn}{Q}}. \tag{3}$$

After this, we can resume processing on a per-subcarrier basis and recover the received signal at subcarrier q as

$$r_q = \mathbf{h}_q^{\intercal} \mathbf{y}_q + \mathbf{v}_q \tag{4}$$

where $\mathbf{h}_q \in \mathbb{C}^{M\times 1}$ is the channel vector at subcarrier q and $\mathbf{v}_q$ is additive white Gaussian noise (AWGN). The full channel and precoding weights across all subcarriers are defined as $\mathbf{H}, \mathbf{W} \in \mathbb{C}^{Q\times M}$.

The wireless channel $\mathbf{h}_q$ is modeled by one of three models. *1)* For extreme frequency-selective fading, the channel at each subcarrier q is modeled as $[\mathbf{h}_q]_m \sim \mathcal{CN}(0,1)$, where the channel at each carrier is independently and identically distributed (i.i.d.). Hence, there is no correlation in the frequency domain. *2)* Frequency flat channels are modeled as $\mathbf{h}_q = \mathbf{h}\ \forall q$, with $[\mathbf{h}]_m \sim \mathcal{CN}(0,1)$. *3)* Channels with a controllable amount of correlation in the frequency domain are obtained using a Rayleigh tapped delay line (TDL) model. The time domain channel impulse response between antenna m and the user is modeled as $\mathbf{h}_m^{(t)} = {\mathbf{C}_\mathbf{h}}^{1/2} \mathbf{g}_m$, with $\mathbf{g}_m \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_L)$. Here $\mathbf{h}_m^{(t)} \in \mathbb{C}^{L\times 1}$ consists of L uncorrelated taps with an exponential decaying power delay profile (PDP), hence the covariance matrix is $\mathbf{C}_\mathbf{h} = \operatorname{diag}\left(\lambda_h^0, \cdots, \lambda_h^{L-1}\right)$ and $\lambda_h^l = e^{-\beta l}$, where β controls the amount of frequency selective fading (high β results in flatter channels, while a low β results in more frequency selectivity). The PDP of tap l is defined as $\mathbb{E}\left(\left|h_l^{(t)}\right|^2\right) = \lambda_h^l$. Additionally, to normalize the power, we set $\operatorname{tr}(\mathbf{C}_\mathbf{h}) = 1$. To get the channel frequency response, the DFT is applied as $\mathbf{h}_m = \mathbf{F}\mathbf{\Sigma}\mathbf{h}_m^{(t)}$, where $\mathbf{F} \in \mathbb{C}^{Q\times Q}$ is the DFT matrix and $\mathbf{\Sigma} = \left[\ \mathbf{I}_L\ \mathbf{0}_{(Q-L)\times L}\ \right]^{\intercal}$ is a zero padding matrix, which appends $Q-L$ zeros to $\mathbf{h}^{(t)}$.

B. Achievable Rate

The Bussgang decomposition at the receiver is used on a per subcarrier basis to obtain the signal-to-noise-and-distortion ratio (SNDR). This is done by decomposing the received signal at subcarrier q as $r_q = B_q s_q + d_q + v_q$, where $s_q \sim \mathcal{CN}(0,1)$ is the intended user symbol at subcarrier q, d_q captures the nonlinear distortion at subcarrier q and v_q is AWGN with variance σ_v^2. The Bussgang gain at subcarrier q is $B_q = \mathbb{E}\left(r_q s_q^*\right) / \mathbb{E}\left(s_q s_q^*\right)$ [10]. Due to the Bussgang decomposition, the distortion d_q, noise v_q and intended signal s_q are uncorrelated, hence the distortion term is $\mathbb{E}\left(|d_q|^2\right) = \mathbb{E}\left(|r_q|^2\right) - |B_q|^2 p_q - \sigma_v^2$, where $p_q = \mathbb{E}\left(s_q s_q^*\right) = 1$. Consequently, the SNDR at subcarrier q is

$$\mathrm{SNDR}_q = \frac{|B_q|^2 p_q}{\mathbb{E}\left(|d_q|^2\right) + \sigma_v^2}. \tag{5}$$

An achievable rate at subcarrier q, R_q, i.e., a lower bound on the capacity, is computed by considering jointly Gaussian distributed noise and distortion, which is independent from the data symbols. This can be seen as a worst case, resulting in

$$R_q = \log_2(1 + \mathrm{SNDR}_q). \tag{6}$$

Additionally, the average rate across all subcarriers is defined as $R_{\mathrm{avg}} = \frac{1}{Q} \sum_{q=0}^{Q-1} R_q$.

C. Loss Function

The loss function used to train the NN-precoder is based on maximizing the average achievable rate across all subcarriers, subject to a power constraint. This is formulated as

$$\max_{\mathbf{W}} R_{\mathrm{avg}}(\mathbf{H}),\ \text{s.t. } p_q \le \frac{P_T}{Q} \tag{7}$$

where $p_q = \sum_{m=0}^{M-1} \mathbb{E}\left(|x_{m,q}|^2\right)$ is the power at carrier q. The total power is defined as $P_T = \sum_{q=0}^{Q-1} p_q$. In this work, the total power is constrained to $P_T = MQ$, which is achieved by setting the per-subcarrier power to $p_q = M$. To achieve this, the precoding vector at subcarrier q is normalized as

$$\mathbf{w}_{q,\mathrm{norm}} = \frac{\sqrt{p_q}}{\|\mathbf{w}_q\|} \mathbf{w}_q. \tag{8}$$

Given that the per-carrier power is set to $p_q = M$, the average power at the input of each PA is $p_{in} = p_q/M = 1$. Next to this, we note that the average input back-off (IBO) over all PAs is defined as $\mathrm{IBO} = p_{in}/p_{sat}$. Given that $p_{in} = 1$ is fixed by the power constraint, we can vary the IBO by varying p_{sat}.

III. GNN Based-precoding

In this work, GNN-based precoding is preferred over other NN architectures due to its permutation equivariance properties, which are beneficial for the precoding task as outlined in [6]. The GNN, $f(\cdot;\boldsymbol{\theta})$, can be seen as a NN that learns a mapping from the channel vector to the precoding vector over a graph, where $\boldsymbol{\theta}$ are the learnable parameters. The graph considered in this work has a node for each transmit antenna and a node for the user. Additionally, each transmit antenna is connected to the user node through an edge, as seen in Fig. 1. The graph is defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \mathcal{V}_M \cup \{u\}$ is the set of nodes, with $\mathcal{V}_M$ being the set of antenna nodes and u being the user node, and $\mathcal{E}$ is the set of edges, which contains all edges between the antenna and user node i.e., $(m, u) \in \mathcal{E}\ \forall m \in \mathcal{V}_M$. The graph is undirected, meaning that $(m, u) \in \mathcal{E} \leftrightarrow (u, m) \in \mathcal{E}$, as is the physical channel between the user and antenna. A general GNN performs I message-passing iterations/layers, which update the hidden representations of the graph edges. For the graph considered in this work, at layer i, a hidden representation is updated for each edge, denoted by $\mathbf{z}_{(m,u)}^{(i)} \in \mathbb{R}^{d_i}$. A single GNN layer that updates the representation of edge (m, u) is defined as

$$\mathbf{z}_{(m,u)}^{(i+1)} = \sigma\left(\mathbf{W}_{\text{edge}}^{(i)}\mathbf{z}_{(m,u)}^{(i)} + \mathbf{W}_{u}^{(i)}\mathbf{m}_{\mathcal{N}(u)}^{(i)}\right) \tag{9}$$

where $\mathbf{z}_{(m,u)}^{(i+1)}$ denotes the hidden representation of the edge (m, u) at layer/iteration $i+1$ of the GNN, $\sigma(\cdot)$ is a non-linear activation function and the matrices $\mathbf{W}_{\text{edge}}^{(i)}, \mathbf{W}_{u}^{(i)} \in \mathbb{R}^{d_{i+1}\times d_i}$ are learnt using a form of stochastic gradient descent. Note that the dimensions of these matrices, d_{i+1} and d_i, determine the number of features in layer $i+1$ and i respectively. The aggregation operation is defined as

$$\mathbf{m}_{\mathcal{N}(u)}^{(i)} = \frac{1}{M}\sum_{m'\in\mathcal{N}(u)} \mathbf{z}_{(m',u)}^{(i)}. \tag{10}$$

Depending on the setting (single-carrier/multi-carrier) the input and output edge features of the GNN differ, as outlined in the next subsections.

A. Single-carrier Case

In our previous work [6], GNN-based distortion-aware precoding was developed for the frequency-flat/single-carrier case. Even though this GNN is trained in the single-carrier case, it can still be applied in a multi-carrier system. For this, two options are available. A first option is to apply the GNN-precoder independently at each subcarrier, to obtain all precoding vectors $\mathbf{w}_q = f(\mathbf{h}_q;\boldsymbol{\theta})\ \forall q$. Hence, the GNN is applied Q times. This ignores the inter-carrier interference due to the non-linear amplifiers. This method is referred to as GNN_1. A second option, referred to as $\text{GNN}_{q_{\text{mid}}}$, is to compute the precoding vector only for the middle carrier using the GNN: $\mathbf{w}_{q_{\text{mid}}} = f(\mathbf{h}_{q_{\text{mid}}};\boldsymbol{\theta})$, where $q_{\text{mid}} = \lfloor\frac{Q-1}{2}\rfloor$. This precoding vector is then applied at all subcarriers. This method is preferred when the channel is fully flat across all subcarriers, as the complexity is reduced by a factor Q as compared to the previous method. In each case, the input features for each edge are the wireless channel between the antenna and the user node at subcarrier q. The output features for each edge are the precoding coefficients. Hence, the input and output for edge (m, u) of the network are defined as $\mathbf{z}_{(m,u)}^{(0)} = \left[\Re\{[\mathbf{h}_q]_m\}, \Im\{[\mathbf{h}_q]_m\}\right]^\intercal \in \mathbb{R}^{2\times 1}, \mathbf{z}_{(m,u)}^{(I-1)} = \left[\Re\{[\hat{\mathbf{w}}_q]_m\}, \Im\{[\hat{\mathbf{w}}_q]_m\}\right]^\intercal \in \mathbb{R}^{2\times 1}$. Given that in the single-carrier case we only consider a single (flat) carrier, which is represented by one complex coefficient, the size of the input and output features are $d_0 = d_{L-1} = 2$, i.e., the real and imaginary parts of the channel and precoding coefficients.

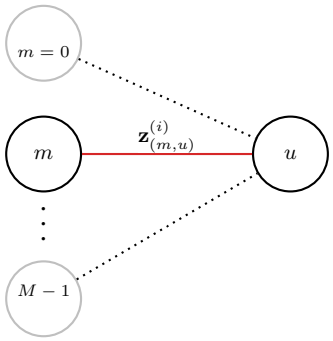

Fig. 1: Graph representing the wireless system. Antenna nodes are on the left, the user node is on the right. Each edge contains an edge feature $\mathbf{z}_{(m,u)}^{(i)}$.

B. Multi-carrier Case

In the multi-carrier case we consider the full channel matrix $\mathbf{H} \in \mathbb{C}^{Q\times M}$, which consists of Q subcarriers, each with its channel vector $\mathbf{h}_q \in \mathbb{C}^{M\times 1}$. In the multi-carrier case, the function to learn is $\mathbf{W} = f(\mathbf{H};\boldsymbol{\theta})$, where $\mathbf{W} \in \mathbb{C}^{Q\times M}$ is a matrix combining the precoding vectors of all subcarriers. This joint precoding over all subcarriers is denoted as GNN_Q. In this case, the input features for edge (m, u) are defined as $\mathbf{z}_{(m,u)}^{(0)} = [\Re\{\mathbf{h}_m\}, \Im\{\mathbf{h}_m\}]^\intercal \in \mathbb{R}^{2Q\times 1}$, where the vector $\mathbf{h}_m \in \mathbb{C}^{Q\times 1}$ contains the channel coefficients for all subcarriers between antenna m and the user. Similarly the output features at edge (m, u) are defined as $\mathbf{z}_{(m,u)}^{(I-1)} = [\Re\{\mathbf{w}_m\}, \Im\{\mathbf{w}_m\}]^\intercal \in \mathbb{R}^{2Q\times 1}$, where the vector $\mathbf{w}_m \in \mathbb{C}^{Q\times 1}$ contains the precoding coefficients for all subcarriers between antenna m and the user. Consequently, the size of the input and output features is $d_0 = d_{I-1} = 2Q$, i.e., the real and imaginary parts of the channel at each subcarrier. Comparing this to the single-carrier case, it can be seen that the input and output feature sizes are scaled by Q.

IV. Inference Complexity

A. Classic Precoding

As a benchmark, we consider MRT precoding. This computes the complex conjugate of the channel vector for each subcarrier, which results in QM floating point operations (FLOPs). Note that the normalization of the precoding vectors is not taken into account in the complexity analysis, as this is the same for all precoder types.

TABLE I: Overview of the different precoders, including feature sizes and complexity.

Precoder	Precoding Function	$d_0 = d_{I-1}$	d_h	Complexity [MFLOPs]
GNN_Q	$f(\mathbf{H};\boldsymbol{\theta})$	$2Q$	256	54.53
GNN_1	$f(\mathbf{h}_q;\boldsymbol{\theta})\ \forall q$	2	64	101.72
$\mathrm{GNN}_{q_{mid}}$	$f(\mathbf{h}_{q_{mid}};\boldsymbol{\theta})$	2	64	3.18
MRT	$(\sqrt{p_q}/\|\mathbf{h}_q\|)\mathbf{h}_q^*\ \forall q$	NA	NA	10^{-3}

B. GNN Single-carrier

In this section, we compute the complexity of the GNN that is trained on a single carrier and then is applied on each carrier independently. At inference time, the GNN is executed for each subcarrier, i.e., Q times. First, we consider a single layer of the GNN which has an input feature size of d_{in} and an output feature size of d_{out}. For this, (9) has to be computed M times (i.e., for each edge), while (10) needs to be computed only once. This leads to the following number of FLOPs

$$\mathcal{O}^{(\text{flops})}_{d_{in},d_{out}} = \underbrace{M d_{out}(4d_{in}-1)}_{(9)} + \underbrace{M d_{in}}_{(10)}. \tag{11}$$

When considering a GNN with I layers and a hidden feature size of d_h we have the following structure: one input layer with $d_{in} = 2$ and $d_{out} = d_h$, $I-2$ hidden layers with $d_{in} = d_{out} = d_h$ and one output layer with $d_{in} = d_h$ and $d_{out} = 2$. When applying this GNN to all Q subcarriers, this results in a total complexity of

$$\begin{aligned}\mathcal{O}^{(flops)}_{GNN_1} &= Q\left(\mathcal{O}^{(flops)}_{2,d_h} + (I-2)\mathcal{O}^{(flops)}_{d_h,d_h} + \mathcal{O}^{(flops)}_{d_h,2}\right)\\ &= 4MQ\left(4d_h + (I-2)d_h^2\right)\end{aligned} \tag{12}$$

Note that the above complexity is for the GNN trained for the single-carrier case which is then applied independently at each subcarrier, this is denoted as GNN_1. Alternatively, we defined $\mathrm{GNN}_{q_{mid}}$ which computes the precoder at the middle carrier and applies it at each subcarrier. Hence, the complexity of $\mathrm{GNN}_{q_{mid}}$ is Q times lower as compared to (12).

C. GNN Multi-carrier

In the multi-carrier case we can reuse the computational complexity of one layer of the GNN as computed in (11). The GNN for joint precoding has the following structure: one input layer with $d_{in} = 2Q$ and $d_{out} = d_h$, $I-2$ hidden layers with $d_{in} = d_{out} = d_h$ and one output layer with $d_{in} = d_h$ and $d_{out} = 2Q$. Given that the GNN in the multi-carrier case jointly computes the precoders across all subcarriers the GNN only needs to be executed one time (as compared to Q times for the single-carrier case). This results in the following computational complexity

$$\begin{aligned}\mathcal{O}^{(flops)}_{GNN_Q} &= \mathcal{O}^{(flops)}_{2Q,d_h} + (I-2)\mathcal{O}^{(flops)}_{d_h,d_h} + \mathcal{O}^{(flops)}_{d_h,2Q}\\ &= 4M\left(4Qd_h + (I-2)d_h^2\right)\end{aligned} \tag{13}$$

The total complexity for each precoder, considering the parameters used in the simulations in Section VI are given in Table I.

V. Rate/Complexity Trade-off Classification

As will be illustrated in Section VI, in most scenarios, joint precoding across all carriers achieves the highest achievable rate (except when the channel is fully flat). However, it holds a high computational cost as compared to MRT and $\mathrm{GNN}_{q_{mid}}$. Hence, a trade-off has to be made. Is a small increase in achievable rate worth a major increase in complexity? To formalize this question, a classifier is proposed that selects the appropriate precoding technique based on the current channel. The main aim of this classifier is to provide a low-complexity decision to determine the appropriate precoder based on the current channel conditions.

The labels to train this classifier are obtained based on the following rate/complexity trade-off. Let's consider the set of the class indices of the classifier as $\mathcal{C} = \{0,1,2,3\}$. The class indices can then be mapped to the appropriate precoder as

$$C : \mathcal{C} \to \{\text{MRT}, \mathrm{GNN}_1, \mathrm{GNN}_{q_{\text{mid}}}, \mathrm{GNN}_Q\} \tag{14}$$

where C denotes a label mapping function from index to label. To determine the label of the classifier based on the current channel $\mathbf{H}$, the following objective function is used

$$c^*(\mathbf{H}) = \arg\max_{c\in\mathcal{C}} \left(R_c(\mathbf{H}) - \alpha\mathcal{O}_c\right) \tag{15}$$

where $c^*(\mathbf{H})$ is the selected label, R_c denotes the achievable rate of precoder c, $\mathcal{O}_c$ the complexity of precoder c and $\alpha \geq 0$ is a scalar complexity regularization coefficient that balances the trade-off between maximizing the achievable rate and minimizing the computational complexity. In this work $\alpha = 1.5 \times 10^{-8}$ is selected as a good trade-off between complexity and rate. An overview of the objective function for the different precoders for this α value is illustrated in Fig. 2. From this figure, it is clear that GNN_1 is never selected as it is penalized by its high complexity. Next to this we see that, generally speaking, for low β values MRT is set as the label, for high β values $\mathrm{GNN}_{q_{mid}}$ is selected and in the middle range ($\beta \sim \pm 1$) GNN_Q is set as the label. The selection of the precoding method depends strongly on the degree of frequency selectivity of the channel. As such, we propose the use of a convolutional neural network (CNN), where learnable filters are convolved over the antenna and subcarrier dimension of the channel $\mathbf{H} \in \mathbb{C}^{Q\times M}$, to perform the classification. This is further detailed in Section VI

VI. Simulation Results

For the simulations, $M = 32$ transmit antennas and $Q = 32$ subcarriers are considered. The polynomial PA coefficients from [6] are used for an IBO value of -3 dB. The linear PA

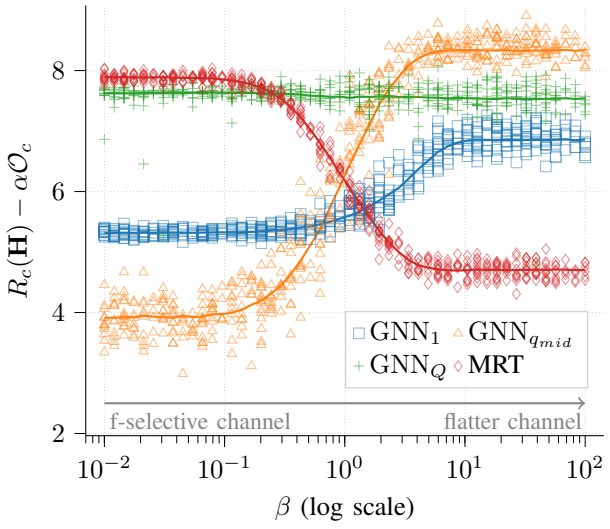

Fig. 2: Classifier objective functions for different precoders vs β (PDP decay factor controlling frequency selectivity) on 500 channel realizations from the test set, with $\text{SNR}_q = 20$ dB and $\alpha = 1.5 \times 10^{-8}$.

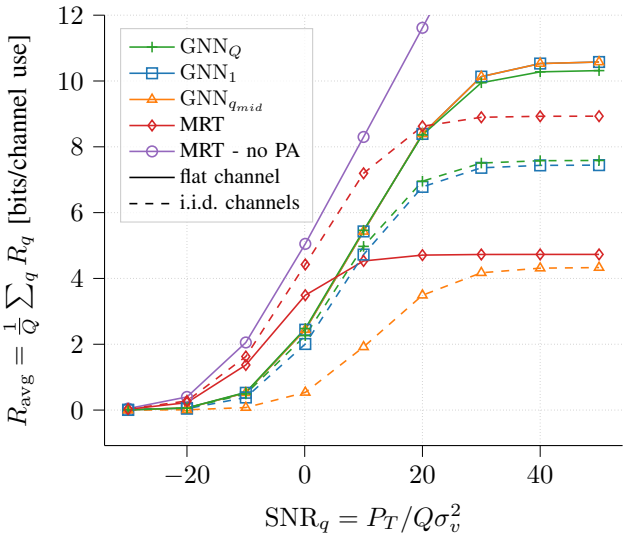

Fig. 3: Rate vs SNR_q averaged over the channels of the test set. For $M = 32, Q = 32, \text{SNR}_q = 20$ dB, on strongly frequency selective/i.i.d. (dashed) and frequency flat channels (full).

gain is set to one. The dataset for the GNN-based precoders consists of a training set of 500.000 channels, a validation set of 1000 channels and a test set of 10.000 channels. For GNN$_Q$ the channels are generated according to the TDL channel model with a β value that is sampled from 50 possible values logarithmically spaced between 10^{-2} and 10^2. The models are trained for 30 epochs, using the Adam optimizer with a learning rate of 0.5×10^{-3} and batch size of 64. Training is done at $\text{SNR}_q = P_T/Q\sigma_v^2 = 20$ dB. GNN$_1$ and GNN$_{q_{mid}}$ consist of $I = 8$ layers with a hidden feature size of $d_h = 64$. GNN$_Q$ consists of $I = 8$ layers with a hidden feature size of $d_h = 256$. For both models, a leaky ReLu activation is used. Additionally, the output of the model is normalized according to (8) in order to respect the power constraint. The classification model consists of two convolution layers with 16 kernels of size 3×3, each followed by a batch normalization and a ReLu activation. These layers are followed by global pooling and a fully connected layer of 32 neurons. The dataset for the classifier consists of a training set of 50.000 channels, a validation set of 5.000 channels and a test set of 10.000 channels. The classifier is trained for 50 epochs, using the Adam optimizer with a learning rate of 10^{-4} and batch size of 32.

A. GNN-based Precoding

To evaluate the performance of the different precoders, different channel conditions are considered. First, two extreme cases are considered in Fig. 3, where fully frequency flat and fully frequency uncorrelated (i.i.d.) channels are considered. When i.i.d. channels are considered, MRT has the best performance as compared to the GNN-based precoders. When the channels are i.i.d., the distortion is more uniformly distributed in space, due to the addition of many uncorrelated components (subcarriers), hence the distortion does not receive a considerable beamforming gain. This is a similar phenomenon to the case where many users are present [11]. In this case, the non-linear distortion is more spatially spread out, due to the addition of many uncorrelated components (users). When frequency flat channels are considered, the distortion is beamformed into the user direction, leading to a lower performance of MRT. Under these channel conditions the GNN-based precoders can achieve a considerable increase in rate of 5.4 bits/ channel realization at an $\text{SNR}_q = 30$ dB. This rate increase is achieved by beamforming the distortion in non-user directions [6]. Notably, we see that under flat channel conditions, all three GNN schemes achieve approximately the same performance.

Second, the TDL channel model is considered, where the amount of frequency selectivity is tuned by changing β. We note that for this channel model, by considering a limited number of taps $L = 10$ (with $L \ll Q$), the frequency domain channel is never fully uncorrelated. In Fig. 4, the rate is illustrated for varying degrees of frequency selectivity (β). When β is large, the channel can be considered flat, resulting in the GNN-based precoders outperforming MRT. As β decreases, the channel becomes more frequency selective, leading to a decreased performance of GNN$_1$ and GNN$_{q_{mid}}$ as these NNs are trained for the single-carrier/flat case. Additionally, as β decreases, the performance of MRT improves, due to the non-linear distortion becoming more spatially spread out. By jointly precoding all subcarriers, as done by GNN$_Q$, we observe that the performance is maintained, even for small values of β. Notably, even for very low β values GNN$_Q$ outperforms MRT.

B. Rate-Complexity Classification

The labels used to train the classifier are illustrated in Fig. 2 and correspond to the precoder that maximizes the objective function. After training, the classifier achieved 95.75% accu-

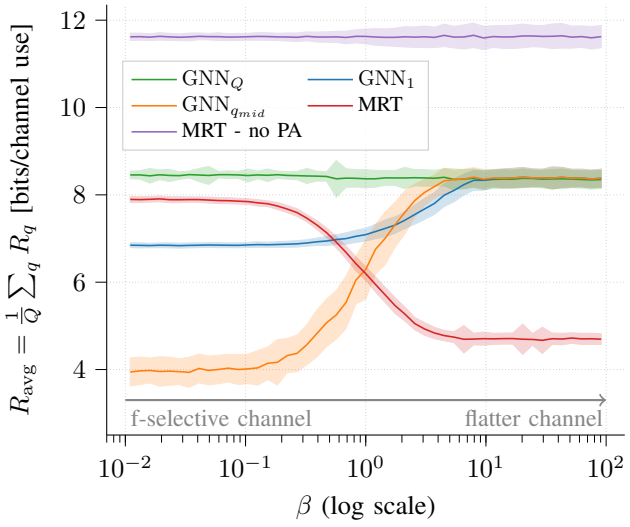

Fig. 4: Rate vs β with ± 1 Standard Deviation averaged over the channels of the test set. For $M = 32$, $Q = 32, \text{SNR}_q = 20$ dB on a channel with varying degrees of frequency selectivity.

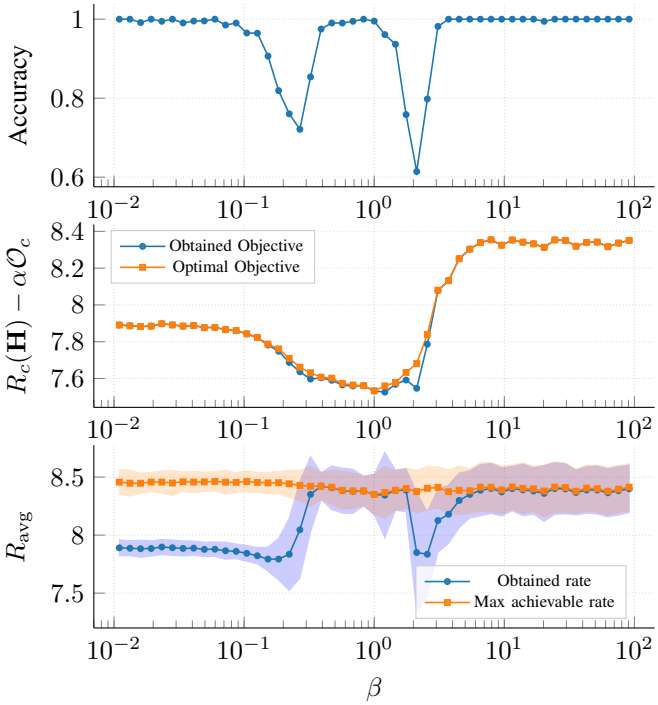

Fig. 5: Accuracy (top), objective function (middle), and average rate (bottom) of the classifier for varying β, which controls the frequency-selectivity.

racy on the test set. In Fig. 5, the top plot illustrates how this accuracy is distributed over the varying channel conditions (β). The accuracy is nearly 100% except in two transition regions, where it drops to about 60%. These regions correspond to changes in the optimal precoder: from MRT to GNN$_Q$, and from GNN$_Q$ to GNN$_{q_{mid}}$. While the hard accuracy metric drops sharply here, the middle plot shows that the actual objective function is barely affected, and the classifier still selects near-optimal precoding. Accordingly, the bottom plot in Fig. 5 confirms that the resulting rate reduction is small, at most 0.75 bits per channel realization. This small reduction in rate is the outcome of the rate/complexity trade-off considered by the objective function.

VII. Conclusion

This work addressed the challenge of distortion-aware precoding in multi-carrier large-antenna systems with non-linear PAs. We showed that distortion-aware per-subcarrier precoders, while effective in single-carrier or frequency-flat channels, suffer performance loss in the presence of PA-induced inter-carrier interference, particularly in highly frequency-selective channels. To mitigate this, we developed a GNN precoder that jointly optimizes across all subcarriers. To reduce the complexity increase related to this joint approach, a CNN-based rate/complexity classifier was proposed to dynamically select between MRT, single-carrier GNN, and joint GNN precoding depending on the current channel conditions. Simulation results confirm that this adaptive strategy balances achievable rate and complexity.

References

[1] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *IEEE wireless communications*, vol. 18, no. 5, pp. 40–49, 2011.

[2] S. K. Mohammed and E. G. Larsson, "Constant-Envelope Multi-User Precoding for Frequency-Selective Massive MIMO Systems," *IEEE Wireless Communications Letters*, vol. 2, no. 5, pp. 547–550, 2013.

[3] P. P. Ann and R. Jose, "Comparison of PAPR reduction techniques in OFDM systems," in *2016 International Conference on Communication and Electronics Systems (ICCES)*, 2016, pp. 1–5.

[4] S. R. Aghdam, S. Jacobsson, and T. Eriksson, "Distortion-Aware Linear Precoding for Millimeter-Wave Multiuser MISO Downlink," in *2019 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2019, pp. 1–6.

[5] F. Rottenberg, G. Callebaut, and L. Van der Perre, "The Z3RO Family of Precoders Cancelling Nonlinear Power Amplification Distortion in Large Array Systems," *IEEE Transactions on Wireless Communications*, vol. 22, no. 3, pp. 2036–2047, 2023.

[6] T. Feys, L. Van der Perre, and F. Rottenberg, "Toward Energy-Efficient Massive MIMO: Graph Neural Network Precoding for Mitigating Non-Linear PA Distortion," *IEEE Transactions on Cognitive Communications and Networking*, vol. 11, no. 1, pp. 184–201, 2025.

[7] A. M. Elbir, K. V. Mishra, M. R. B. Shankar, and B. Ottersten, "A Family of Deep Learning Architectures for Channel Estimation and Hybrid Beamforming in Multi-Carrier mm-Wave Massive MIMO," *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 2, pp. 642–656, 2022.

[8] S. Jacobsson, G. Durisi, M. Coldrey, and C. Studer, "Linear Precoding With Low-Resolution DACs for Massive MU-MIMO-OFDM Downlink," *IEEE Transactions on Wireless Communications*, vol. 18, no. 3, pp. 1595–1609, 2019.

[9] T. Schenk, *RF Imperfections in High-rate Wireless Systems: Impact and Digital Compensation*. Dordrecht: Springer Netherlands, 2008.

[10] O. T. Demir and E. Bjornson, "The Bussgang Decomposition of Nonlinear Systems: Basic Theory and MIMO Extensions [Lecture Notes]," *IEEE Signal Processing Magazine*, vol. 38, no. 1, pp. 131–136, 2021.

[11] C. Mollen, U. Gustavsson, T. Eriksson, and E. G. Larsson, "Spatial Characteristics of Distortion Radiated From Antenna Arrays With Transceiver Nonlinearities," *IEEE Transactions on Wireless Communications*, vol. 17, no. 10, pp. 6663–6679, 2018.