Virtual DPD Neural Network Predistortion for OFDM-based MU-Massive MIMO

Chance Tarver*, Alexios Balatsoukas-Stimming†, Christoph Studer‡, and Joseph R. Cavallaro*
*Department of Electrical and Computer Engineering, Rice University,
†Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands
‡Department of Information Technology and Electrical Engineering, ETH Zürich, Zürich, Switzerland

***Abstract*—The nonlinearities of power amplifiers in massive MIMO arrays introduce unwanted spectral regrowth, which is typically avoided via digital predistortion at each amplifier. However, as the number of base station antennas scales up, so does the computational burden of per-antenna linearization. This work introduces a neural-network virtual digital predistortion (vDPD) scheme that operates before the linear precoder for OFDM-based massive MU-MIMO systems. By applying predistortion before the precoder, complexity scales primarily with the number of users. We can achieve comparable linearization along the user beams by training our neural network based on the memory polynomial, predistortion-per-antenna approach. We verify our algorithm through an exhaustive simulator that includes high-order amplifier nonlinearities, memory effects, and variance across the amplifier models.**

I. Introduction

In the last ten years, massive multiple-input multiple-output (MIMO) has gone from initial conception to commercial deployment. While deployments are underway, there are still practical challenges that have not been fully considered. Of note, the linearization of large antenna arrays poses a significant computational burden for base stations [1]. This topic has received recent interest in the literature with discussions of the effects [2, 3], possible technologies [1], and some possible digital predistortion (DPD) solutions [4–6]. A detailed analysis on the effect of nonlinearities is studied in [2]. However, it is assumed that each antenna uses exactly the same power amplifier (PA), which is not realistic due to process variations during the manufacturing of the PAs. The work of [3] considers a more general case with unique PA models for each antenna. However, throughout the analysis, memory effects are not fully considered, which are certainly present in practical scenarios [7]. This neglect of memory effects is potentially problematic for the wide bandwidths considered in beyond-5G technologies.

Energy efficiency is a key design goal for MIMO arrays. PAs typically consume the most power in transmitters, so their efficiency is critical. In [1], Doherty PAs are considered a good candidate for large antenna arrays due to their energy efficiency. However, they are also known for being highly nonlinear, meaning that DPD will be necessary [7]. However, the common DPD approach of using an inverse model of a PA scales poorly to large antenna arrays, as one inverse model is required for each antenna, and few works offer solutions to reduce the computational burden for the linearization of massive MIMO arrays. In [4], complexity is reduced for linearizing a MIMO array by using a lower-complexity decorrelation-based DPD method. However, the complexity of this method still scales with the number of antennas. When performing DPD, it is clear that high-complexity solutions result in additional power consumption in the form of additional digital signal processing (DSP) blocks on FPGA implementations or additional area on ASIC implementations. This complexity threatens the energy efficiency goals of massive MIMO.

One significant finding throughout the recent literature is that, in beamformed massive MIMO, the majority of the adjacent channel power (ACP) due to PA nonlinearities follows the main beams [2, 3]. Based on this, multiple works have presented a beam-oriented DPD [5, 6]. In these works, the main objective is to linearize along the main beam by only considering the nonlinearity experienced by the intended user. However, these works do not consider the case of orthogonal frequency-division multiplexing (OFDM) modulation found in 5G, which poses a new challenge to these previous methods in that the modulation and precoding are done in the frequency domain. In contrast, the DPD methods are developed for the time domain. Moreover, these works do not consider the case of multiple users (MU-MIMO), where the linear combination of the data caused by the linear precoding leads to intermodulation of user data and new spurious out-of-band (OOB) beams.

In this work, we expand the beam-oriented DPD concept and introduce a virtual DPD scheme (vDPD) that operates on the primary beams where the distortion is strongest as well as the secondary spurious beams that may appear in some systems. In particular, we utilize a neural network (NN) to perform a nearly identical DPD calculation to that done by a typical memory polynomial (MP)-based DPD. This vDPD scheme operates before the precoder so that complexity can be reduced.

The primary contributions of this work are as follows: 1) We provide a mathematical derivation for the expected directions of spurious beams for a typical array and precoding scheme. 2) We introduce an NN-based DPD scheme before the precoder, vDPD. 3) We show with simulation results that our vDPD method can achieve sufficient performance in all directions from the array. 4) We provide a complexity comparison showing a complexity reduction under steady-state operation.

This work was supported in part by the US NSF under grant CNS-2016727.

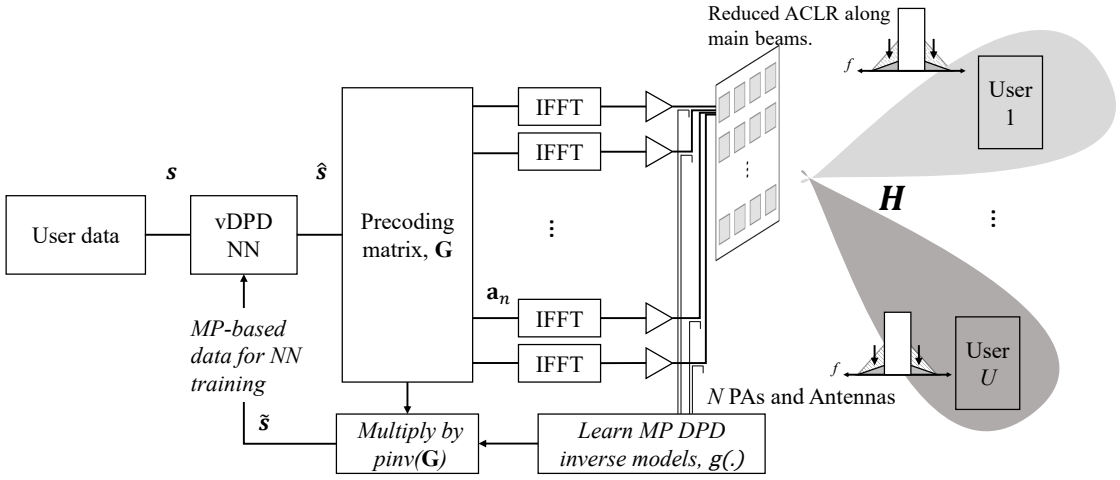


Figure 1. Block diagram for the vDPD system. User data is updated before the precoder with the goal of applying an equivalent operation to the standard DPD at each antenna.

II. System Model and Algorithm

A. System Model

We consider a fully digital, multi-user (MU) massive MIMO system with U single antenna users and N PAs and antenna RF units at the base station (BS). Without loss of generality, we restrict the presentation below to one OFDM symbol. A symbol of data to the users is represented by the vector $\mathbf{s}_w \in \mathcal{O}^U$, where w indexes the OFDM tones from 1 to W and $\mathcal{O}$ represents the set of complex-valued constellation points. Pulse shaping is applied via the inclusion of guard-band subcarriers that are normally empty.

Linear precoding is applied separately to each OFDM tone, generating W vectors $\mathbf{x}_w \in \mathbb{C}^N$ with $\mathbf{x}_w = \mathbf{G}_w\mathbf{s}_w$. Here, $\mathbf{G}_w \in \mathbb{C}^{N\times U}$ is the precoding matrix such as zero forcing (ZF) or maximum ratio transmission (MRT). Each vector is remapped to contain all the tones per antenna, $[\mathbf{x}_1, \ldots, \mathbf{x}_W] = [\mathbf{a}_1, \ldots, \mathbf{a}_N]^T$, where each $\mathbf{a}_n$ is a W-dimensional vector containing all tones for antenna port $n \in \{1, \ldots, N\}$. At this point, the data is converted from the frequency domain to the time domain via the inverse discrete Fourier transform (IDFT), which is typically calculated via an inverse fast Fourier transform (IFFT). The data is reorganized to be serial instead of parallel, and a cyclic prefix is added. In many systems, windowing is also applied between symbol boundaries to improve the spectral shaping. We express this time-domain representation for each antenna as the vector $\mathbf{u}_n$. This vector is upconverted to an RF frequency where it is transmitted through a PA with nonlinear function $f_n(\cdot)$. The time-domain data for each antenna is given as $\hat{\mathbf{u}}_n = f_n(\mathbf{u}_n)$, equivalently expressed as a discrete-time signal, $\hat{u}[i] = [\mathbf{u}_n]_i$. The frequency-domain equivalent is given as $\hat{\mathbf{x}}_n$.

In OFDM systems, the channel is usually modeled in the frequency-domain for each tone w as, $y_w = \mathbf{h}_w\hat{\mathbf{x}}_w + n_w$, where y_w denotes the received data for OFDM tone w and $\mathbf{h}_w$ is the $1 \times N$ channel vector, and n_w is a Gaussian random noise term. The user received signal can be remapped to $[y_1, \ldots, y_W] = \mathbf{b}$ to represent a W dimensional vector of all tones received at the user. The time-domain user-received signal is given as $\mathbf{v}$.

B. MIMO and Spurious Beams

The question of how the unintended ACP is radiated in MIMO systems has seen recent attention in the literature [2, 3, 6, 8]. However, the analysis is often only done to some mathematical conclusions without providing any practical answer as to which direction do the spurious beams point in. Ultimately, the answer to this question will depend on the array geometry considered and the precoding scheme. However, we will answer this question for a half-wavelength spaced uniform linear array (ULA) with MRT beamforming in a line-of-sight (LoS) channel.

To build our analysis, we begin with a two-tone signal similar to the work done in [3]. Consider two users with incident angles at θ_1 and θ_2. Assuming a planar wave model, the channel vector to the i-th user can be written as,

$$[\mathbf{h_i}]_n = \left[e^{-j\pi n \cos\theta_i}\right], \tag{1}$$

where $n = 0, ..., N-1$ indexes the base station antenna [9]. The MRT precoder, $\mathbf{p}_i$, is then formulated via the complex conjugate of Eq. (1).

Consider the case of transmitting a tone to each user at baseband frequencies ω_1 and ω_2. After precoding, the composite signal at each PA input would be,

$$x_n(t) = e^{j(\omega_1 t + \phi_{1,n})} + e^{j(\omega_2 t + \phi_{2,n})}, \tag{2}$$

where the phase of each tone is given as

$$\phi_{i,n} = \pi n \cos\theta_i. \tag{3}$$

While in Section IV we consider high-order MP PA models with memory for simulation, in our analysis we consider

the following memoryless third-order model for mathematic tractability,

$$f_n(x) = x + \alpha_n x|x|^2, \quad (6)$$

where α_n is the complex coefficient specific to the model PA n. After Eq. (2) is substituted into Eq. (4), we get tl output of each PA as

$$\begin{aligned} y_n(t) = &(1+3\alpha_n)\, e^{j(\omega_1 t+\phi_{1,n})} \quad (7)\\ &+ (1+3\alpha_n)\, e^{j(\omega_2 t+\phi_{2,n})} \quad (8)\\ &+ \alpha_n e^{j((2\omega_1-\omega_2)t+(2\phi_{1,n}-\phi_{2,n}))} \quad (9)\\ &+ \alpha_n e^{j((2\omega_2-\omega_1)t+(2\phi_{2,n}-\phi_{1,n}))}. \quad (10) \end{aligned}$$

The goal is then to find the physical directions with respect to the array, $\hat{\theta}$, that correspond with the spurious beams caused by the terms from (7) and (8). Without loss of generality, we focus on the term from (7) and look for the angle that maximizes the array response,

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \sum_{n=0}^{N-1} \alpha_n e^{j(2\omega_1-\omega_2)t+(2\phi_{1,n}-\phi_{2,n}-\pi n\cos\theta)}. \quad (9)$$

In cases where all PAs have identical phases on α_n, the coefficient can be eliminated as the phase shift would be common across all elements in the array. Otherwise, as they become more randomly distributed, the less coherently the intermodulations combine [3]. To solve the above, we note that the beamforming does not depend on t and set it to zero. We note that the sum will be maximized when, if possible, all the exponential terms are cophased. We choose to force the phase of each exponential term to be zero and solve for the θ that allows for that. With the above assumptions and replacing the ϕ terms back with their exact values from Eq. (3), we arrive at the direction of interest for the term from (7) as,

$$\hat{\theta} = \cos^{-1}\left(2\cos\theta_1 - \cos\theta_2\right). \quad (10)$$

While the two-tone analysis is simple, the conclusions scale-up similarly to OFDM [3]. When scaling up to more than two users, the spurious beams will scale up with every pair and triple of users. The main idea is to find the number of terms similar to (7) and (8) as more users are added. From analysis similar to deriving (7) and (8), it can be shown that there will be a third-order intermodulation (IM3) term for each pair and triple of user signals. In the case of the previous two-tone, a three-user system will have terms with phases of the form as $\phi_{1,n} + \phi_{2,n} - \phi_{3,n}$. This tripple can be arranged in three unique ways, and, similarly, each pair of terms can be arranged in two ways. Combining these ideas, the upper bound on the total number of spurious beams created by the IM3 between user streams is given as, $n_{\text{spurious beams}} = 2\binom{U}{2} + 3\binom{U}{3}$. This is considered an upper bound since for certain user angles, for example, when users are regularly spaced at some angle, these spurious beams may be in the same direction and appear as a single beam. There will be more terms and spurious beams for higher-order nonlinearities, but the third-order intermodulations are typically the highest magnitude and, hence, are the primary concern.

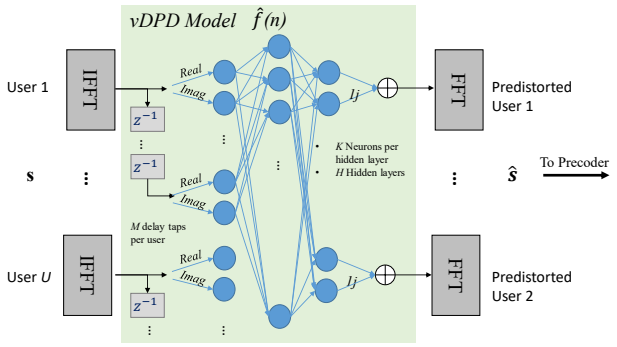


Figure 2. vDPD Structure. The neural network models the function that is needed to predistort each user stream.

III. Virtual DPD NN Algorithm

The system architecture of our proposed vDPD scheme for MU-massive MIMO is illustrated in Fig. 1. Our method's main idea is to train a NN to emulate the response of the widely adopted MP-based solution in the lower dimensional space before the precoder.

A. Conventional Per-Antenna DPD

Extending the conventional DPD work to apply to multi-antenna systems can be done by applying a DPD block per antenna. In Eq. (11) a MP predistorter for PA n, $g_n(.)$, is shown with nonlinearity order P and M memory taps:

$$\hat{u}^{(n)}[i] = \sum_{p=1}^{P}\sum_{m=0}^{M} \beta_{p,m}^{(n)} u^{(n)}[i-m]\left|u^{(n)}[i-m]\right|^{p-1}. \quad (11)$$

An indirect learning architecture (ILA) [10] is often used to set up a least-squares problem to solve for the predistorter coefficients. An optimal set of coefficients for antenna port n, $\beta^{(n)}$, is found so that the cascade of the predistorter block and the power amplifier is linear [11]. While this method is commonplace and performs well, it scales poorly in many-antenna scenarios in that it has to be done for every PA. It operates on each time-domain, digital-baseband signal with no knowledge of the full system or the modulation scheme.

B. Novel vDPD Neural Net

To correct the nonlinearity experienced by the user, we aim to learn a single NN-based *virtual DPD*. The vDPD structure is shown in Fig. 2. Each user stream is converted to the time domain, where it will go through an M-tap delay line. A real value decomposition occurs on the complex samples so that the real and imaginary parts of each go into their own input neurons. The vDPD scheme includes H hidden layers, each with K neurons. The hidden layer neurons utilize a nonlinear activation function such as the ReLU. At the final layer, there is a neuron for the real and imaginary components of each of the U streams. The time-domain outputs are converted back to the frequency domain as $\hat{\mathbf{s}}$ before being sent to the precoder. A linear bypass can also be utilized to pass the input data directly to the output neurons so that the NN only learns the nonlinearities [12].

1) NN Training: To train any NN, the construction of training data can be critical. We utilize the MP-based DPD function of each PA in Section III-A, which is widely adopted for linearization. The MP output for each symbol, $\tilde{\mathbf{u}}$ is converted back to the frequency domain for each subcarrier to create $\tilde{\mathbf{x}}$. To create the $U \times 1$ vector of NN output training data for each subcarrier, we multiply by the psuedo-inverse of each precoder, $\tilde{\mathbf{s}}_{\mathbf{w}} = \mathbf{G}_w^{\dagger} \tilde{x}_w$. The NN can then be trained using standard optimizers such as gradient descent and a mean-squared error loss.

While the training complexity is high, it is run relatively infrequently. The MP-DPDs that are used to create the training data do not need to be relearned often since they are based on the PAs, whose models remain relatively consistent for a given temperature. The NN will need to be retrained for a new precoder, but this can be done in few epochs by using the previous weights as starting values and simply learning an update.

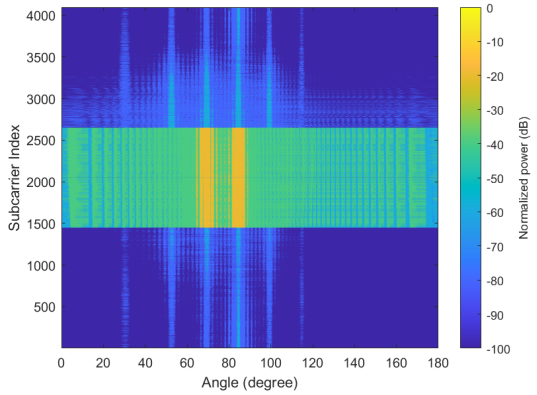
2) NN Application: Given data for a user, we can perform NN inference to obtain the predistorted version of the user data, $\hat{\mathbf{s}}$ for each subcarrier. The modulation steps outlined in Section II-A and shown in Fig. 1 are then performed as usual.

In this work, we assume perfect channel state information (CSI). Practically, it would not be possible to directly measure CSI on the guard-band subcarriers as the users do not transmit pilots on these subcarriers. However, it may be possible to extrapolate these subcarriers by extending known CSI or applying interpolation-based techniques common for OFDM denoising [13].

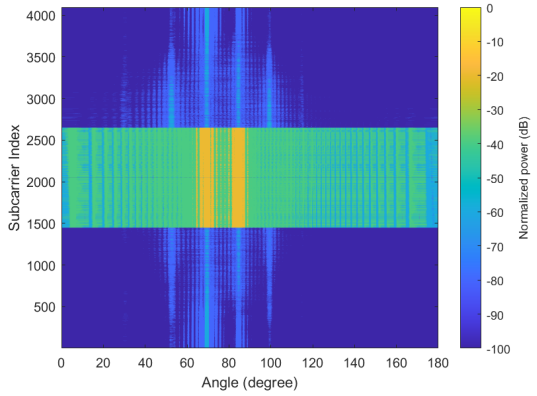
IV. DPD Simulation Results

In this section, we explore massive MIMO simulations under the presence of unique MPs with $P = 7$ and $M = 4$. PA coefficients for each antenna are drawn from a normal distribution with a 10% variance around a PA model collected from the WARPv3 SDR platform. The combination of using a measurement-based PA model with memory effects and variance between all PAs in the array contributes to making this simulation realistic. $D = 1200$ OFDM data subcarriers out of $W = 4096$ total subcarriers are beamformed by a zero-forcing precoder to a user at 70° and a second user 85°, 400 meters from the uniform linear array operating at 3.5 GHz. We assume that the base station has full CSI and that there is a line-of-sight channel. The QuaDRiGa channel model [14] software environment is used to create the simulation topology.

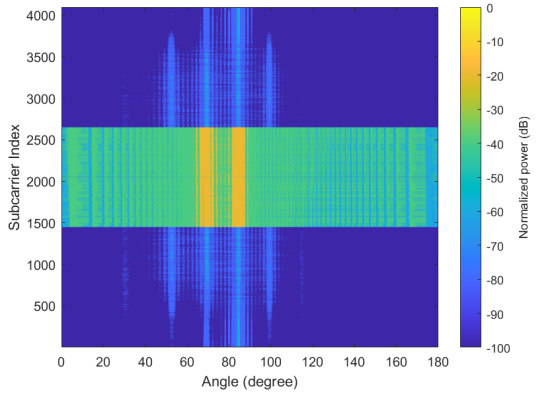
We consider the radiated power as a function of angle and subcarrier index in Fig. 3. For this plot, vertical cross-sections are equivalent to a power spectral density (PSD), and horizontal cross-sections are equivalent to a beamforming plot. Fig. 4 shows the beampatterns for the out-of-band adjacent channel. We train a MP-per-antenna DPD with $P = 7, M = 4$ to create the vDPD training data. Its performance is shown in Fig. 3c. The $K = 15, M = 4, H = 1$ vDPD is then trained. In Fig. 3a where no DPD is applied, it is shown that the ACP is dominant in the user's direction while there are minor nonlinearities appearing in other spatial directions. This finding



(a) Without DPD



(b) After vDPD



(c) After MP

Figure 3. Radiated power as a function of angle and subcarrier for users placed at 70° and 85° from a ULA with $N = 64$ antennas. The subcarrier index corresponds to each 15 kHz bin used in the modulation. The center 1200 subcarriers correspond to in-band data. The lower subcarriers correspond to the first lower frequency OOB channel while the upper subcarriers correspond to the 1st upper frequency adjacent channel.

is consistent with other works in the literature [2, 3]. Fig. 3b shows the radiated power after our vDPD. Here, the OOB emissions in the direction of the user have been reduced. There are still minor nonlinearities appearing in other directions, but the vDPD scheme also reduces these spurious OOB beams.

In Fig. 4 we examine the OOB beam pattern by plotting the array response for the upper adjacent channel from Fig. 3.

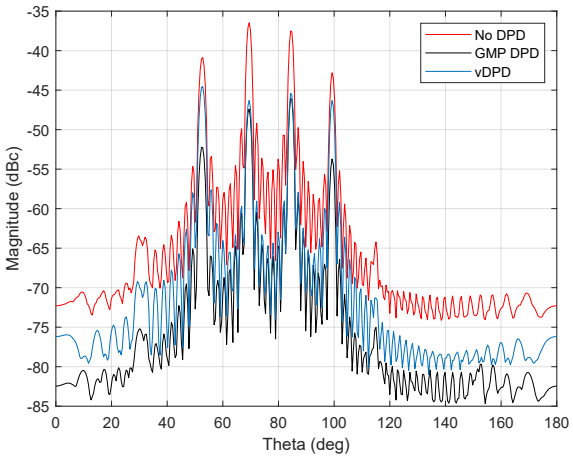


Figure 4. Maximum OOB radiation relative to the main carrier in the upper adjacent channel (dBc) for each angle before and after vDPD. By applying the vDPD scheme, we are able to reduce the radiation along the main beams (70° and 85°) and spurious beams.

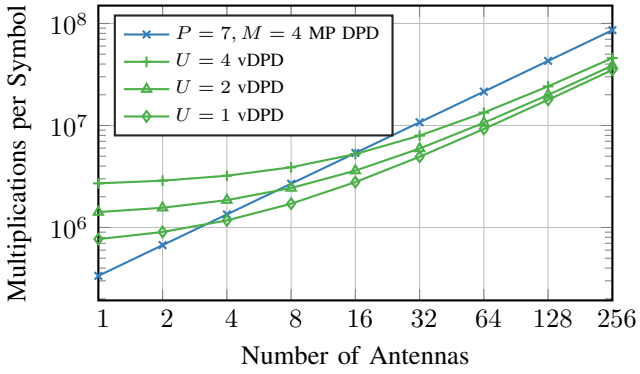


Figure 5. Multiplications per OFDM symbol versus the number of antennas. The NN is fixed to $M = 4, K = 15, H = 1$.

With PAs in the system, we start with the case of no DPD shown in red. The vDPD method reduces the power in the user directions by as much as 10 dB with 4 dB reduction in other directions. The in-beam directions have performance matching the MP training data, while the spurious directions have a minor ACP reduction.

a) Algorithm Complexity: Using commonplace approaches such as the MP DPD will require linearizing each PA individually as outlined in Section III-A. When considering the large number of antennas considered in 5G and beyond, the complexity can quickly become prohibitive. The main advantage of our proposed approach is that, while the MP-per-PA approach scales with the number of antennas, the vDPD uses a small NN. In Fig. 5 we plot the complexity of the MP-per-antenna DPD and the vDPD versus the number of transmit antennas N. Here, we fix the memory to $M = 4$, for all systems, and we consider the case where there are 1200 data subcarriers. The signal is upsampled to 4096 samples per OFDM symbol for both DPDs to be over 3x upsampling. The MP-per-antenna DPD, shown in blue, increases linearly as each new antenna requires a new MP. The vDPD, shown in green, requires only one NN for all cases. We consider a fixed number of hidden layers, $H = 1$, and number of neurons, $K = 15$, per hidden layer. The input and output layers scale with the number of users. While the NN size is fixed in this figure, complexity increases with the number of antennas due to multiplications associated with the additional precoding of guard-band subcarriers. At $N = 32$, the required number of multiplications for our vDPD is clearly below the case of applying a DPD per PA. In the $N = 64$ case with $U = 2$, corresponding to the simulation results from Fig. 4, the $K = 15$ NN vDPD is able to provide a 50.6% reduction in complexity.

V. Conclusions

In this work, we introduced a novel vDPD scheme targeted at massive MU-MIMO systems using OFDM modulation. By predistorting before the precoding, we effectively linearize the beam to each user. Since most OOB energy follows the main beam, linearizing the beam provides sufficient reduction in ACP for the whole system. Compared to performing DPD on all antennas, vDPD provides similar performance at a fraction of the complexity, which can translate to more energy-efficient massive MIMO implementations.

References

[1] W. Chen, G. Lv, X. Liu, D. Wang, and F. M. Ghannouchi, "Doherty PAs for 5G massive MIMO: Energy-efficient integrated DPA MMICs for sub-6-GHz and mm-wave 5G massive MIMO systems," *IEEE Microw. Mag.*, vol. 21, no. 5, 2020.

[2] C. Mollén, U. Gustavsson, T. Eriksson, and E. G. Larsson, "Spatial characteristics of distortion radiated from antenna arrays with transceiver nonlinearities," *IEEE Trans. on Wireless Commun.*, vol. 17, no. 10, 2018.

[3] L. Anttila, A. Brihuega, and M. Valkama, "On antenna array out-of-band emissions," *IEEE Wireless Commun. Letters*, vol. 8, no. 6, 2019.

[4] M. Abdelaziz, L. Anttila, and M. Valkama, "Reduced-complexity digital predistortion for massive MIMO," in *2017 IEEE Int. Conf. on Acoustics, Speech and Signal Process.*

[5] X. Liu, Q. Zhang, W. Chen, H. Feng, L. Chen, F. M. Ghannouchi, and Z. Feng, "Beam-oriented digital predistortion for 5G massive MIMO hybrid beamforming transmitters," *IEEE Trans. on Microw. Theory Techn.*, vol. 66, no. 7, 2018.

[6] M. Abdelaziz, L. Anttila, A. Brihuega, F. Tufvesson, and M. Valkama, "Digital predistortion for hybrid MIMO transmitters," *IEEE J. of Sel. Topics Signal Process.*, vol. 12, no. 3, 2018.

[7] F. M. Ghannouchi and O. Hammi, "Behavioral modeling and predistortion," *IEEE Microw. Mag.*, vol. 10, no. 7, 2009.

[8] S. R. Aghdam, S. Jacobsson, U. Gustavsson, G. Durisi, C. Studer, and T. Eriksson, "Distortion-aware linear precoding for massive MIMO downlink systems with nonlinear power amplifiers," *CoRR*, vol. abs/2012.13337, 2020.

[9] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge, UK ; New York: Cambridge University Press, 2005.

[10] C. Eun and E. J. Powers, "A new Volterra predistorter based on the indirect learning architecture," *IEEE Trans. on Signal Process.*, vol. 45, no. 1, 1997.

[11] L. Ding, G. T. Zhou, D. R. Morgan, Z. Ma, J. S. Kenney, J. Kim, and C. R. Giardina, "A robust digital baseband predistorter constructed using memory polynomials," *IEEE Trans. Commun*, vol. 52, no. 1, 2004.

[12] C. Tarver, A. Balatsoukas-Stimming, and J. R. Cavallaro, "Design and implementation of a neural network based predistorter for enhanced mobile broadband," in *2019 IEEE Int. Workshop on Signal Process. Syst.*

[13] S. Haene, A. Burg, N. Felber, and W. Fichtner, "OFDM channel estimation algorithm and ASIC implementation," in *2008 4th European Conf. on Circuits and Syst. for Commun.*

[14] S. Jaeckel, L. Raschkowski, K. Börner, and L. Thiele, "QuaDRiGa: A 3-D multi-cell channel model with time evolution for enabling virtual field trials," *IEEE Trans. on Antennas and Propag.*, vol. 62, no. 6, 2014.