Neural Network Based Digital Predistortion for Active Antenna Arrays Under Load Modulation

Alberto Brihuega, Lauri Anttila, and Mikko Valkama

***Abstract*—In this letter, we propose an efficient solution to linearize mmWave active antenna array transmitters that suffer from beam-dependent load modulation. We consider a dense neural network that is capable of modeling the correlation between the nonlinear distortion characteristics among different beams. This allows to provide consistently good linearization regardless of the beamforming direction, avoiding thus the necessity of executing continuous digital predistortion parameter learning. The proposed solution is validated, conforming to 5G New Radio transmit signal quality requirements, with extensive over-the-air RF measurements utilizing a state-of-the-art 64-element active antenna array operating at 28 GHz carrier frequency.**

***Index Terms*—5G New Radio, digital predistortion, load modulation, mmWave, nonlinear distortion, neural networks.**

I. Introduction

In order to increase the power-efficiency of active antenna arrays operating at mmWaves, specifically the so-called frequency range 2 (FR2), different polynomial-based digital predistortion (DPD) solutions have been proposed [1]–[8]. In these works, a replica of the main-beam signal is utilized for DPD learning, as most of the nonlinear distortion is radiated towards this direction [9]. One key challenge in this context is the so-called beam-dependent load modulation [10]. Due to the high level of integration of mmWave array transmitters, where isolators between antennas and PA units are preferably avoided [10], the interactions between neighboring antennas cause the PA's output port impedance, and consequently also their nonlinear characteristics, to change with the steering angle [3], [4], [10]. As a result, a given DPD system that was initially trained considering a specific beam-direction or nonlinear conditions, may no longer be adequate for a new beam-direction [3], [4]. In 5G New Radio (NR), the beam direction can be updated at a millisecond level, thus this imposes very tight time constraints on the DPD parameter estimation – potentially implying even continuous adaptation. Due to the large signal bandwidths in FR2 systems, real time adaptation may be challenging. This constraint greatly contrasts with common adaptation times in ordinary single- or few-antenna systems, where the DPD system only needs to track long-term variations, e.g., due to temperature drifts.

As an alternative to polynomial-based DPD solutions, neural networks (NNs) can offer effective linearization solutions due to their excellent nonlinear modelling capabilities [11]–[13]. Different NN topologies have been considered in the DPD context, while most recent ones are based on real input data, where the I and Q components along with the envelope of the signal are considered for training the NN [11], [12]. Additionally, [12] augments the NN input layer so that it is capable of modeling possible I/Q imbalance and crosstalk effects. As in physical systems, the nonlinearities primarily act on the envelope of the signal, thus [13] proposed to only utilize the envelope of the signal as input to the NN, along with a phase recovery stage at the last hidden layer to obtain the complex baseband signal. However, similar to traditional polynomial-based DPD solutions, regular NNs should be re-trained as the beamforming direction changes – something that is very challenging since generally NNs require even longer training times compared to polynomial-based solutions.

Manuscript received May 8, 2020; revised May 23, 2020; accepted June 12, 2020.

The research work leading to these results was supported by the Academy of Finland (under the projects #304147, #301820, and #319994), Nokia Bell Labs, and the Tampere University Doctoral School.

A. Brihuega, L. Anttila, and M. Valkama are with the Department of Electrical Engineering, Tampere University, Tampere, Finland.

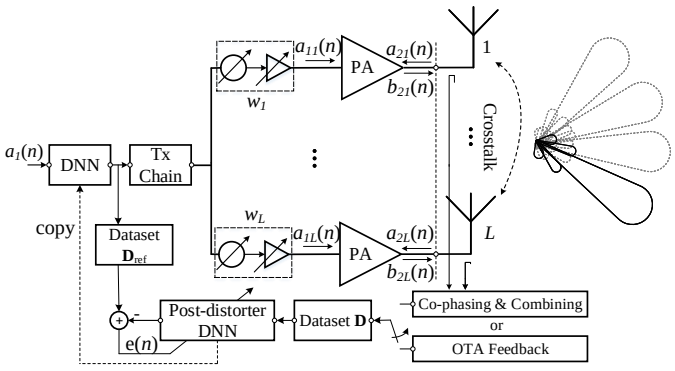

Fig. 1. Considered active array transmitter with DNN-based DPD.

In this letter, we propose to utilize a dense NN (DNN) that is trained with a data set that contains signals from different beamforming directions. Despite the antenna array exhibiting distinct nonlinear characteristics for every beam direction, the resulting radiated distortions at different directions are mutually correlated. This can thus be exploited by the DNN, if properly trained, so that consistently good linearization is provided across the whole range of beamforming directions. Once the DNN is trained properly, only occasional offline re-learning is needed to adapt to long-term variations, similar to ordinary DPD systems in single or few-antenna transmitters. The proposed concept is formulated in Section II while RF measurement results are provided in Section III.

II. Proposed Dense Neural Network Based DPD

A. Nonlinear Active Array Model

We consider an active antenna array transmitter that is depicted in Fig. 1. Due to the mutual coupling between the antenna elements, the coupled waves drive the output ports of the PAs, which results in a dynamic variation of the PAs' output port impedances which in turn modifies their

nonlinear characteristics. Under these circumstances, dual-input behavioral models [14], [15] are commonly adopted to accurately model the PAs' behaviour. Given the considered transmitter architecture, such model can be reduced to classical single-input behavioral model as described in [3], yielding

$$\begin{aligned} b_{2i}(n) = &\sum_{m_1=0}^{M_1}\sum_{p=0}^{(P_1-1)/2} \alpha_{m_1}^{(2p+1)} w_i |w_i|^{2p} a_1(n-m_1) \\ &\times |a_1(n-m_1)|^{2p} \\ &+ \sum_{m_2=0}^{M_2} \beta_{m_2}^{0} f_i(n-m_2) \star a_1(n-m_2) \\ &+ \sum_{m_3=0}^{M_3}\sum_{m_4=0}^{M_4}\sum_{p=1}^{(P_2-1)/2} \beta_{m_4 m_3}^{2p+1} |w_i|^{2p} \\ &\times f_i(n-m_3) \star a_1(n-m_3) |a_1(n-m_4)|^{2p} \\ &+ \sum_{m_5=0}^{M_5}\sum_{m_6=0}^{M_6}\sum_{p=1}^{(P_3-1)/2} \zeta_{m_6 m_5}^{2p+1} w_i^2 |w_i|^{p-1} \\ &\times f_i^*(n-m_5) \star a_1^*(n-m_5) (a_1(n-m_6))^2 \\ &\times |a_1(n-m_6)|^{2(p-1)}, \end{aligned} \tag{1}$$

where P_1, P_2, P_3 denote polynomial orders, $M_1, \ldots, M_6$ designate memory depths of the model, while $\alpha_{m_1}^{(2p+1)}$, $\beta_{m_2}^{0}$, $\beta_{m_4 m_3}^{2p+1}$ and $\zeta_{m_6 m_5}^{2p+1}$ are the model coefficients or free parameters. Additionally, $f_i(n) = \sum_{l=1}^{L} w_l \lambda_{il}(n) \star \mu_l(n)$ where $\lambda_{il}(n)$ is the filter impulse response that models the crosstalk from the lth to the ith antenna, $\mu_l(n)$ is an impulse response that models the linear distortion in the lth antenna/PA branch, $\star$ denotes the convolution operator, while L is the total number of antennas.

Despite the beam-dependent load-modulation impacting the exact nonlinear characteristics of the antenna array, formally expressed through the filter impulse response $f_i(n)$, there is strong mutual correlation between the nonlinear distortion radiated towards different beamforming directions. This is visible from (1), where the model parameters defining the nonlinear behavior of the PAs, i.e., $\alpha_{m_1}^{(2p+1)}$, $\beta_{m_2}^{0}$, $\beta_{m_4 m_3}^{2p+1}$ and $\zeta_{m_6 m_5}^{2p+1}$ do not depend on the beamforming direction. Based on this, we argue that a single DPD system should be capable of providing sufficient linearization regardless of the beamforming direction, but only if it is trained so that the overall nonlinear distortion – i.e., distortion radiated across the range of considered beamforming directions – is considered. This will be verified with measurements in Section III.

B. Proposed Multi-beam Oriented DNN-based DPD

In order to effectively linearize the antenna array for every beamforming direction with a single DPD unit, we consider a feedforward fully connected DNN with one hidden layer with G neurons. The DNN architecture is depicted in Fig. 2, where the transmit signal I and Q components, denoted as $I_{\text{in}}(n)$ and $Q_{\text{in}}(n)$, respectively, as well as the pth order envelopes $|a_1(n)|^p$ for $p = 1, 2, \ldots P_{\text{DNN}}$ are used as the inputs. In order to model wideband memory effects, also the delayed replicas up to memory depth of M_{DNN} are adopted. Furthermore, the hidden layer is assumed to utilize the hyperbolic tangent sigmoid activation function [11].

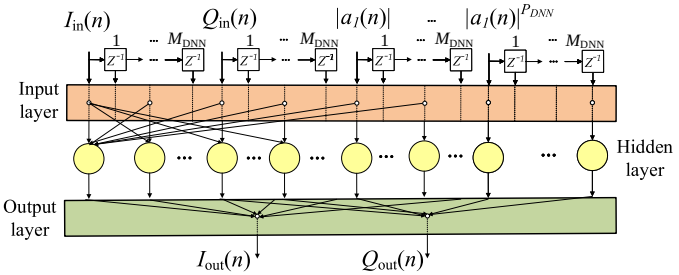

Fig. 2. Block diagram of the DNN DPD system.

Algorithm 1 DNN DPD Training Procedure

1: Generate the dataset $\mathbf{D}$ based on (2) and $\mathbf{D}_{\text{ref}}$ accordingly
2: Train the post-distorter DNN with $\mathbf{D}$ and $\mathbf{D}_{\text{ref}}$
3: Copy the DNN as pre-distorter and transmit
4: Repeat from step 1 until convergence

In order to train the NN, we consider an indirect learning architecture (ILA), where a DNN-based post-inverse is trained by considering replicas of the far-field beamformed signals for different beam directions. Then, the weights of the DNN post-distorter are copied as the actual DNN DPD, as illustrated in Fig. 1, while the process is iterated until convergence. More specifically, within one ILA iteration, the feedback samples of signals belonging to K different main-beam directions are concatenated into a large single vector

$$\mathbf{D} = [\mathbf{r}_{\theta_1}^T \mathbf{r}_{\theta_2}^T \ldots \mathbf{r}_{\theta_K}^T]^T, \tag{2}$$

where the vectors $\mathbf{r}_{\theta_\mathbf{k}}$, $k = 1, 2 \ldots, K$, contain N samples of the signal beamformed towards θ_k. These feedback samples can be obtained e.g., through a hardware-based combiner and an observation receiver [1], or through over-the-air (OTA) feedback from a far-field receiver, with both options being depicted in Fig. 1 [4], [10]. The HW-based combining solution is preferable in the sense that it would also allow to keep track of the long-term variations by occasionally gathering beam-based data during the online data transmission. The reference data for training, denoted as $\mathbf{D}_{\text{ref}}$ is the DPD output as depicted in Fig. 1, and must follow the same structure as that used for $\mathbf{D}$, i.e., it is built by concatenating the signals $a_1(n)$ that reflects the digital transmit waveform behind the feedback signals $\mathbf{r}_{\theta_k}^T$, $k = 1, 2, \cdots, K$.

During the training, the following cost function of the form

$$e = \frac{1}{2B} \sum_{n=1}^{B} \left((I_{\text{out}}(n) - I_{\text{ref}}(n))^2 + (Q_{\text{out}}(n) - Q_{\text{ref}}(n))^2 \right), \tag{3}$$

is calculated for every epoch, with batch size of B, where $I_{\text{ref}}(n)$, $Q_{\text{ref}}(n)$ are the I and Q components of the reference data. Then, the synaptic weights and biases of the different neurons are updated utilizing back-propagation approach and the Levenberg-Marquardt algorithm [16], until convergence. Minimizing e implies minimizing the distortion across the considered range of beam directions. It is noted that, as memory effects are considered in the DNN, the last M_{DNN} samples of $\mathbf{r}_{\theta_k}$ will impact the training of $\mathbf{r}_{\theta_{k+1}}$. However M_{DNN} is generally much smaller than N, thus the impact is negligible. The training procedure is summarized in Algorithm 1.

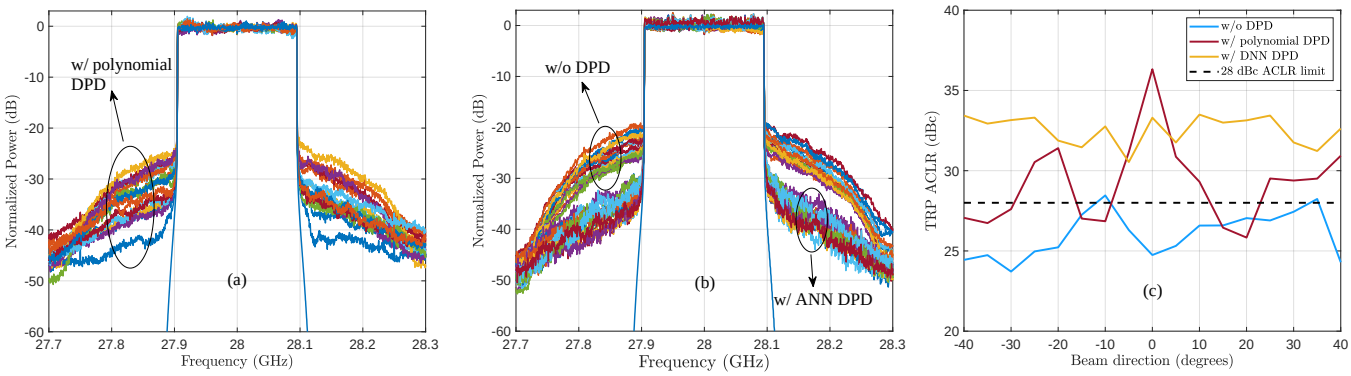

Fig. 3. OTA observed spectra at 28 GHz with (a) the reference PW-CL DPD [4] trained at zero degrees direction, (b) the proposed DNN DPD. The spectra w/o DPD are only shown in (b). In (c), the TRP-based ACLRs as functions of the electrical beam direction are shown. The EIRP is $+42$ dBm.

Fig. 4. mmWave OTA measurement setup.

III. mmWave Over the Air Measurements

A. mmWave OTA Measurement Setup

The measurement setup is depicted in Fig. 4. The M8190A arbitrary waveform generator generates the TX IF signal centered at 3.5 GHz. Two N5183B-MXG signal generators running at 24.5 GHz are used as local oscillator that together with two highly linear T3-1040 mixers, are utilized for up-converting the IF signal to the 28 GHz carrier frequency, at TX, and for downconverting the signal back to IF, at RX. The modulated RF waveform is amplified with one HMC499LC4 and one HMC1131 driver amplifiers that allow to feed the Anokiwave AWMF-0129 active antenna array such that its PAs are driven close to saturation, delivering up to $+42$ dBm EIRP. The OTA transmit signal is captured by a horn-antenna located 1.5 meters apart. After downconversion to IF, the signal is captured by the DSOS804A oscilloscope, and taken to baseband. The received samples are then processed in a host PC running MATLAB. The measurement setup is calibrated to reduce the impact of other HW components, such that the observed signal quality is mainly dictated by the active antenna array. As the considered active array does not facilitate HW-based combining, the true OTA RX signal is deployed for DPD learning.

A 200 MHz 5G NR OFDM waveform with 64-QAM subcarrier modulation, subcarrier spacing of 60 kHz, 3168 active subcarriers, and basic FFT size of 4096 is adopted. With an oversampling factor of 5, the TX and RX sample rate is 1.228 GHz. The sample-level PAPR of the DPD input signal, after iterative clipping and filtering, reads 7 dB when measured at 10^{-4} CCDF point. For training, validation and testing, we consider beamforming directions from -40 to 40 degrees with a resolution of 5 degrees, resulting in $K = 17$, each of them having $N = 8k$ samples, out of which 70% are used for training, 15% for validation and 15% for testing. 4 ILA iterations are considered, each employing between $3 - 30$ epochs containing independent symbol realizations in each block. Memory-depth of $M_{\text{DNN}} = 2$ and envelope orders up to $P_{\text{DNN}} = 5$ are considered in the proposed DNN DPD, with $G = 40$ neurons, while the PW-CL DPD from [4] with the same parameterization as the one used in [4, Section VII] is considered as reference.

B. Measurement Results

Snapshot linearization performance examples in terms of spectra are provided in Fig. 3 (a) and (b). As can be observed, the linearization provided by the PW-CL DPD is good at few beam directions only, close to the learning angle, while then degrades substantially at those directions that exhibit more distinct distortion. On the other hand, the proposed DNN DPD is capable of providing consistently good linearization for the whole considered range of beamforming directions. Fig. 3 (c) then illustrates the linearization performance in terms of the ACLR measured through the total radiated power (TPR) metric defined by 3GPP [17], while it also illustrates the ACLR limit for systems operating at FR2 [17]. As can be observed, the proposed DNN DPD allows to fulfill the TRP ACLR target across the whole range of beamforming directions, at the cost of slightly reducing the maximum linearization performance. The DNN DPD thus avoids the need for beam-level parameter learning, opposed to the PW-CL DPD, whose performance falls clearly below the ACLR target when switching to beam directions further away from the learning beam. Consequently, the proposed DNN DPD stands as an effective solution to linearize active antenna arrays affected by load-modulation.

IV. Conclusions

In this letter, a novel NN based DPD solution and parameter learning approach were proposed to linearize active antenna arrays that are affected by beam-dependent load modulation. The proposed approach avoids the need for continuous beam-based parameter learning in the online operation, while it was shown through RF measurements to be capable of linearizing a state-of-the-art active antenna array regardless of the beamforming direction. Future work considers applying the proposed DPD learning approach to polynomial-based DPD solutions.

REFERENCES

[1] M. Abdelaziz, L. Anttila, A. Brihuega, F. Tufvesson, and M. Valkama, "Digital Predistortion for Hybrid MIMO Transmitters," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 445–454, June 2018.

[2] 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. Microw. Theory Techn.*, vol. 66, no. 7, pp. 3419–3432, July 2018.

[3] E. Ng, Y. Beltagy, G. Scarlato, A. B. Ayed, P. Mitran, and S. Boumaiza, "Digital predistortion of millimeter-wave RF beamforming arrays using low number of steering angle-dependent coefficient sets," *IEEE Trans. Microw. Theory Techn.*, pp. 1–14, 2019.

[4] A. Brihuega, M. Abdelaziz, L. Anttila, M. Turunen, M. Allén, T. Eriksson, and M. Valkama, "Piecewise digital predistortion for mmWave active antenna arrays: Algorithms and measurements," *IEEE Trans. Microw. Theory Techn.*, pp. 1–1, 2020.

[5] C. Yu, J. Jing, H. Shao, Z. H. Jiang, P. Yan, X. Zhu, W. Hong, and A. Zhu, "Full-angle digital predistortion of 5G millimeter-wave massive MIMO transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 7, pp. 2847–2860, July 2019.

[6] X. Liu, W. Chen, L. Chen, F. M. Ghannouchi, and Z. Feng, "Linearization for hybrid beamforming array utilizing embedded over-the-air diversity feedbacks," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 12, pp. 5235–5248, Dec. 2019.

[7] Q. Luo, X. Zhu, C. Yu, and W. Hong, "Single-receiver over-the-air digital predistortion for massive MIMO transmitters with antenna crosstalk," *IEEE Trans. Microw. Theory Techn.*, pp. 1–15, 2019.

[8] A. Brihuega, L. Anttila, M. Abdelaziz, T. Eriksson, F. Tufvesson, and M. Valkama, "Digital predistortion for multiuser hybrid MIMO at mmWaves," *IEEE Trans. Signal Process.*, pp. 1–1, 2020.

[9] L. Anttila, A. Brihuega, and M. Valkama, "On antenna array out-of-band emissions," *IEEE Wireless Commun. Lett.*, vol. 8, no. 6, pp. 1653–1656, Dec. 2019.

[10] C. Fager, T. Eriksson, F. Barradas, K. Hausmair, T. Cunha, and J. C. Pedro, "Linearity and efficiency in 5G transmitters: New techniques for analyzing efficiency, linearity, and linearization in a 5G active antenna transmitter context," *IEEE Microw. Mag.*, vol. 20, no. 5, pp. 35–49, May 2019.

[11] D. Wang, M. Aziz, M. Helaoui, and F. M. Ghannouchi, "Augmented real-valued time-delay neural network for compensation of distortions and impairments in wireless transmitters," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 1, pp. 242–254, Jan 2019.

[12] P. Jaraut, M. Rawat, and F. M. Ghannouchi, "Composite neural network digital predistortion model for joint mitigation of crosstalk, I/Q imbalance, nonlinearity in MIMO transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 11, pp. 5011–5020, 2018.

[13] Y. Zhang, Y. Li, F. Liu, and A. Zhu, "Vector decomposition based time-delay neural network behavioral model for digital predistortion of RF power amplifiers," *IEEE Access*, vol. 7, pp. 91559–91568, 2019.

[14] K. Hausmair, S. Gustafsson, C. Sánchez-Pérez, P. N. Landin, U. Gustavsson, T. Eriksson, and C. Fager, "Prediction of nonlinear distortion in wideband active antenna arrays," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 11, pp. 4550–4563, Nov. 2017.

[15] H. Zargar, A. Banai, and J. C. Pedro, "A new double input-double output complex envelope amplifier behavioral model taking into account source and load mismatch effects," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 2, pp. 766–774, Feb. 2015.

[16] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the Society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.

[17] 3GPP Tech. Spec. 38.104, "NR; Base Station (BS) radio transmission and reception," v15.4.0 (Release 15), Dec. 2018.