Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Highly non-linear and wide-band mmWave active array OTA linearisation using neural network

Jalili, Feridoon; Zhang, Yufeng; Hintsala, Markku; Jensen, Ole Kiel; Chen, Qingyue; Shen, Ming; Pedersen, Gert Frølund

[Link to publication from Aalborg University](#)

The Institution of
Engineering and Technology WILEY

Highly non-linear and wide-band mmWave active array OTA linearisation using neural network

Feridoon Jalili[1] (iD) | Yufeng Zhang[1] | Markku Hintsala[2] | Ole Kiel Jensen[1] |
Qingyue Chen[1] | Ming Shen[1] | Gert Frølund Pedersen[1]

[1]Department of Electronic Systems, Aalborg University, Aalborg, Denmark

[2]Rohde & Schwarz Finland Oy, OY, Finland

Correspondence

Feridoon Jalili, Department of Electronic Systems, Aalborg University, Aalborg 9220, Denmark.
Email: fja@es.aau.dk and mish@es.aau.dk

Funding information

Innovationsfonden

Abstract

This paper proposes a neural network (NN)-based over-the-air (OTA) linearisation technique for a highly non-linear and wide-band mmWave active phased array (APA) transmitter and compares it with the conventional memory polynomial model (MPM)-based technique. The proposed NN effectively learns the distinctive non-linear distortions, which may not easily fit to existing MPM solutions, and can, therefore, successfully cope with the challenges introduced by the high non-linearity and wide bandwidth. The proposed technique has been evaluated using a state-of-the-art 4 × 4 APA operating in highly non-linear regions at 28 GHz with a 100-MHz-wide 3GPP base-station signal as input. Experimental results show the pre-distortion signal generated by the NN exhibits the peak-to-average power ratio (PAPR) much lower than the one generated by MPM and consequently superior linearisation performance in terms of adjacent channel leakage ratio (ACLR) and error vector magnitude (EVM) for high non-linearity cases. Using the proposed NN-based linearisation technique, an improvement of 5-dB ACLR and 7% points in EVM are achieved, which demonstrates the promising potential of this technique for emerging broadband communication systems such as 5G/6G and low Earth orbit (LEO) satellite networks.

KEYWORDS

5G mobile communication, active antenna arrays, learning (artificial intelligence), linearisation techniques, millimetre wave amplifiers, millimetre wave antenna arrays, satellite communication

1 | INTRODUCTION

Recent wireless communication systems operating at mmWave are using active phased array (APA) transmitters together with multi-input multi-output (MIMO) systems to improve the system capacity and data rates of the wireless networks. The traditional linearisation techniques like digital pre-distortion (DPD) that are mainly based on the memory polynomial model (MPM) are facing new challenges [1] such as the following.

- Highly integrated front-ends and a large number of PA's do not allow placement of feedback circuits for each branch so a single-input single-output (SISO) over-the-air (OTA) model is needed.

- mmWave frequencies and wide-band linearisation must be handled.
- The increasing desire for high power efficiency requires linearisation of power-efficient but highly non-linear APAs.

Several solutions using a SISO model and modified DPD algorithms for linearisation have been proposed to combat the above challenges [2–6]. A SISO modelling where the entire transmitter is considered as a two-port system has been presented in [7–10] using an observation receiver in far-field together with using an MPM-based DPD technique for linearisation of the antenna array in the presence of crosstalk. In [11], the potential mismatches between PAs have been

compensated, so that they all exhibit the very same behaviour. By doing so, it is possible to provide linearisation in all directions with a single DPD, in contrast to linearising the main beam only. However, compensating the mismatch requires analogue circuits, which introduces complexity and delay for large arrays and the potential changes in the PAs' behaviours due to crosstalk. In the present work, the reference signal for DPD learning was obtained through measurements from a far-field test receiver placed on the main beam direction and the focus is on the challenges related to high bandwidth and high non-linearity.

For the cases where the enhanced power efficiency is required such as Doherty PAs and envelope-tracking-based transmitters, the amplifier exhibits different behaviour for different power levels. A piecewise model based on a region partition algorithm that takes the actual non-linear characteristics of the device into account was proposed in [10], which gives significantly better linearisation than the general memory polynomial models. However, memory modelling capabilities may be compromised in piecewise models as the different sub-models operate independently, whereas memory effects may involve samples belonging to different sub-regions. A new piecewise model for PAs based on the mixture-of-experts (ME) approach, which builds on a probabilistic model that allows the different sub-models to cooperate, has been presented in [12]. It demonstrates a model that outperforms previous piecewise modelling methods. The ME approach is a promising technique and is a highly valid approach to be compared with the neural network (NN) approach in future work. The challenges as high bandwidth and high non-linearity lead to huge complexity and explosion of MPM-based algorithms. The Volterra series model approach, which is commonly used in MPM approaches, is preferable if the order of the non-linearity is not too high (e.g. third or possibly fifth order) [13]. With very high non-linearity order, the MPM approach is not practical because of the increased complexity and consequent increase of the number of unknown kernel coefficients in the model [14]. Neural networks (NNs) are well known to be able to learn any arbitrary non-linear function according to the universal approximation theorem [15]. Several state-of-the-art linearisation techniques based on NN have recently been introduced. A solution for performance imperfections such as crosstalk, power amplifier (PA) non-linearities along with modulator imperfections like in-phase and quadrature (I/Q) imbalance and DC-offset for a wide-band direct-conversion transmitter has been recently introduced in [16, 17]. A similar approach, where only the magnitude of the input signal undergoes a non-linear operation and the phase information is recovered with a linear weighting operation, has been introduced in [18].

For wideband signals, in particular, the memory effects have a significant impact [19, 20]. To take care of memory effects, two dynamic neural structures have been proposed in the NN literature [21]. In the first structure, recurrent neural networks (RNNs) utilise feed-forward and feedback signal processing. In another structure, a real-valued time-delay neural network (RVTDNN) combines I/Q processing with input time-delay lines to handle memory effects, whereas RNN uses output-to-input time-delay lines. Reference [22] indicates that RVTDNNs offer superior performance and easy baseband implementation when used for inverse modelling of PAs with strong non-linearities and memory effects. For the high non-linearity cases, the model needs a low learning rate during training at the cost of the training time. In the present paper, we are using the so-called batch normalisation (BN) together with the hidden layer in order to use a higher learning rate and reduce the training time. Furthermore, the proposed RVTDNN uses the rectified linear units (ReLU) activation function, which is less computationally expensive than hyperbolic tangent (Tanh) and Sigmoid because it involves simpler mathematical operations [23]. We are proposing an NN using only one hidden layer and a minimum number of neurons to make it comparable with conventional MPM. The proposed RVTDNN is applied to linearise highly non-linear multi-PA devices-under-test (DUTs) such as the active phased array. We are using the proposed NN model for a 5G DUT that includes complex interactions between the PAs in the array, such as load-modulation. Measurements quantifying this impact are included in Section 6. Finally, for the first time, to the best knowledge of the authors, a pre-distortion scheme based on the RVTDNN was validated using a real 5G test-bed environment with a minimum number of neurons and layers together with the ReLU activation function to keep the cost and size of the device during implementation as low as possible.

Figure 1 illustrates the digital pre-distortion concept for the APA based on the equivalent SISO model using the proposed neural network. The mapping relationship between the order of memory depth, the number of hidden layers and the number of neurons to the corresponding required linearity have been analysed. The optimum levels for the parameters in each block have been identified, verified through measurements, and then, the performance and the complexity are compared with the applied MPM-based DPD using the same laboratory setup.

This paper is organised as follows: Section 1 is the introduction. Section 2 describes the MPM-based approach. Section 3 is about the NN linearisation technique. Section 4 explains the NN training and parameter tuning, Section 5 is an investigation of complexity and Section 6 is about the measurement results. A discussion on comparison between measurements results of MPM and NN approaches is included in Section 7, and finally, the conclusion of this work is presented in Section 8.

2 | MPM-BASED APPROACH

The classical approach to modelling the full behaviour of a non-linear device is by the Volterra series, Equation (1), which describes the relation between the output and input signals in discrete time:
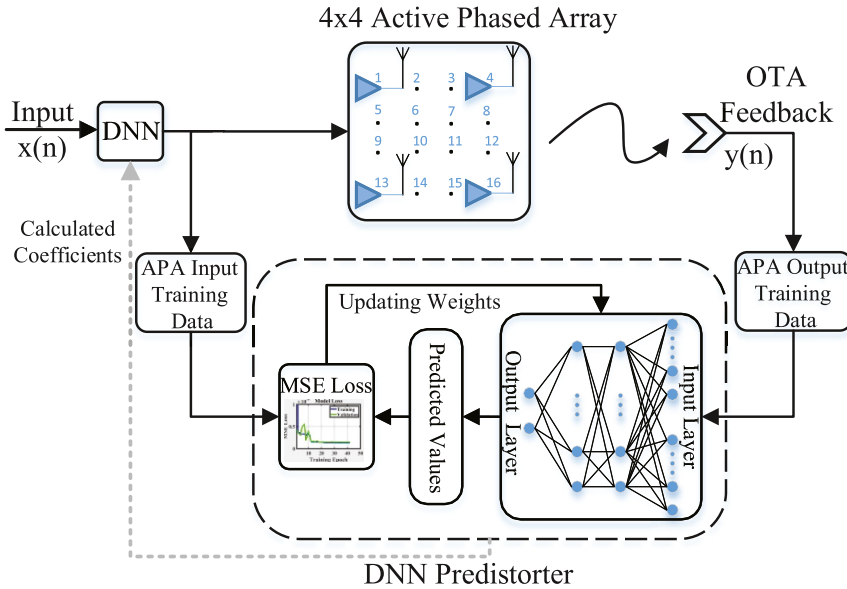
FIGURE 1 Concept illustration of the digital pre-distortion for active phased array (APA) based on the equivalent single-input single-output model using the neural network. DNN, delay neural network

$$
\begin{aligned}
y[n] &= \sum_{k=1}^{K} \sum_{m_1=0}^{M-1} \dots \sum_{m_k=0}^{M-1} h_k[m_1, \dots, m_k] \prod_{j=1}^{k} x[n - m_j] \\
&= \sum_{m_1=0}^{M-1} h_1[m_1] x[n - m_1] \\
&+ \sum_{m_1=0}^{M-1} \sum_{m_2=0}^{M-1} h_2[m_1, m_2] x[n - m_1] x[n - m_2] \\
&+ \dots \\
&+ \sum_{m_1=0}^{M-1} \dots \sum_{m_K=0}^{M-1} h_K[m_1, \dots, m_K] \prod_{j=1}^{K} x[n - m_j],
\end{aligned}
\tag{1}
$$

where K is the order of the non-linearity, M is the memory depth and $h_k(m_1, \dots, m_k)$ are the parameters of the model, which are often referred to as the 'Volterra kernels' in the literature. The nth sample of the input signal $x[n]$ is mixed with the $M - 1$ preceding samples at each of the kth Volterra kernel. In other words, the kth kernel includes all possible combinations of k time shifts of the input signal, which includes all types of memory effects. For this reason, the Volterra series is considered as the most complete model, but the computational complexity of the model is very high [24]. A much less complex model is the MPM, which is widely used for linearisation. Equation (2) represents the applied MPM that is a deviation of the Hammerstein model and has been proven effective for removing non-linearity and memory effect [25]:

$$
y(n) = \sum_{k=1}^{K} \sum_{m=0}^{M} a_{km} x(n - m) |x(n - m)|^{k-1},
\tag{2}
$$

where a_{km} is the 2-D array of filters and power-series coefficients of the active device, K is the non-linearity order of the memory polynomial model and M is the highest memory depth. a_{km} coefficients are the linear weighting of non-linear signals and these coefficients are calculated by using the least-squares type algorithm. The generalised memory polynomial that combines the memory polynomial with cross terms between the signal and lagging and/or leading exponentiated envelope terms is presented in [25]. This model shows a slightly improved linearisation effect but on the cost of complexity that needs to be compared with a more complex neural network model, that is, long short-term memory (LSTM) neural network techniques [26]. In this work, we introduce the comparison between a MPM model based on Equation (2) and a simple neural network model to relax the overall complexity.

In [8], we have provided a detailed insight into the linearisation mechanisms for an APA based on the MPM model. A similar approach has been used for constructing the pre-distorted signals for the different non-linearity cases of the APA in actual work. The same captured input and output I and Q samples are used for both MPM and NN techniques.

3 | NN LINEARISATION TECHNIQUE

3.1 | NN model

The SISO model where the entire transmitter has been considered as a two-port system has been described in [8]. This model uses only one external antenna for observing the combined signal in the far field. Similarly, in the present work, the entire OTA beam-forming setup is considered as a SISO model with the APA as the main source of non-linearity. The NN is used as an inverse system for such a model and it is trained using the measured input and output data. Once the training is completed, the inverse model is used as a pre-distorter for the SISO model, as seen in Figure 2. If the output and the input of the SISO model are set to the I and Q training data, $y(t)$, and true values, $x(t)$, respectively, then the NN needs to be trained to capture the non-linearity of the model by generating the inverse of the non-linearity function, $h(t)$, given in Equation (3):
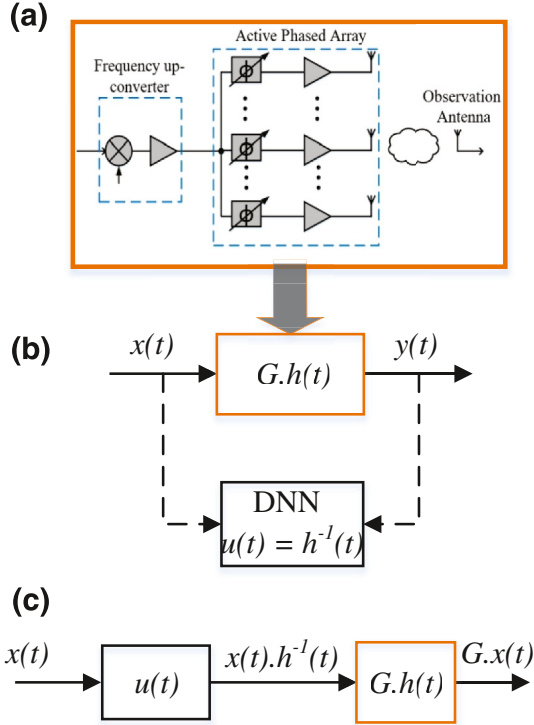
FIGURE 2 Linearisation technique based on a neural network. (a) Beam-forming behavioural single-input single-output (SISO) model; (b) neural network (NN) based on the equivalent SISO model where G is the constant gain and $h(t)$ is the non-linear function; (c) Applying trained NN in pre-distortion and linearisation

$$y(t) = G \cdot h(t) \cdot x(t), \tag{3}$$

where G is the constant gain.

The aim of the training is to calculate the weights such that the NN gradually learns the non-linearity of the SISO model during the training procedure. When the cost is under the specified threshold or no longer converges, then the training step is finished.

After training, the functionality of the NN, denoted as $u(t)$ in Figure 2, is an optimal estimation of $h^{-1}(t)$. Ideally, using the pre-distorted signal, $x(t) \cdot h^{-1}(t)$, as input to the SISO model, the output will be a linearised function defined as $G \cdot x(t)$.

3.2 | NN training

The proposed NN shown in Figure 3 uses a feedforward fully connected (FC) structure. Based on the interconnection pattern or architecture, we can distinguish between feedforward networks (FNNs) and recurrent (or feedback) networks (RNNs) [27]. The feedforward network is considered since it is the most used NN and according to the universal approximation theorem, it can approximate any non-linear function with any desired error [28]. An FC structure in a densely populated NN may increase requirements for hardware resources, but in many applications, the weight of some interconnections can be set to zero without loss of accuracy, which results in sparsely connected layers [27]. The sparse

structure is out of the scope of this work. The input and output data are separated as $y_I[n - M]$, $y_Q[n - M]$, $x_I[n]$ and $x_Q[n]$ where n is the number of the I and Q data used in the training. The wide-band memory effects are modelled by the delayed replica up to memory depth of M. The weights, $\mathbf{W}_{(i)}$, and biases, $\mathbf{B}_{(i)}$ for the vector expressing the relation between input and output of each FC layer is defined as:

$$\mathbf{y}_{(i)} = \mathbf{W}_{(i)} \cdot \mathbf{x}_{(i)} + \mathbf{B}_{(i)}, \tag{4}$$

where i is the ith FC layer. For an input layer of L neurons and output of P neurons, $\mathbf{x}_{(i)}$ is an $L \times 1$ vector, $\mathbf{W}_{(i)}$ is a $P \times L$ matrix and $\mathbf{B_i}$ is a $P \times 1$ vector. Each dense layer, which is defined as (a) in Figure 3, can be described using Equation (4). The weights and biases of each FC layer are distinctive and are optimised by back propagation. The optimisation algorithm used in this work is the adaptive moment estimator (Adam) [29]. It is based on a gradient descent algorithm that gets more computationally efficient by using momentum and randomised batches to avoid local minima. The batch size is the number of training samples used for estimating the error gradient. A batch size, for example, 50, means that 50 samples of the training samples are used for estimating the error gradient before the weights are updated. Another parameter, called training epoch, shows how many passes have been done through the training samples with a randomly selected group of batches. The training procedure is summarised in Table 1.

3.3 | Accelerating NN training

For the models used in high non-linearity cases, the training needs a low learning rate, which, on the other hand, increases the training time. Therefore, in each hidden layer, there is also an accelerator, the so-called BN layer, shown as block (b) in Figure 3. The BN layer allows using much higher learning rates that will accelerate the training and reduce the time cost significantly [30].

The BN layer normalises the mean and variance of the outputs of the dense layer to zero and one and introduces a new mean and variance. The output of the BN layer $\hat{\mathbf{y}}_{(i)}$ is expressed by:

$$\hat{\mathbf{y}}_{(i)} = \gamma \frac{\mathbf{y}_{(i)} - \mathbb{E}[\mathbf{y}_{(i)}]}{\sqrt{\mathrm{Var}[\mathbf{y}_{(i)}] + \epsilon}} + \beta, \tag{5}$$

where γ and β are the new learnable mean and variance parameters and ϵ is a constant parameter to prevent the equation be infinite and is set to 0.001.

3.4 | Activation function

For the NN to be able to fit an arbitrary non-linear function, the non-linearity is introduced in the form of an activation function, which is shown as block (c) in Figure 3. Both Tanh
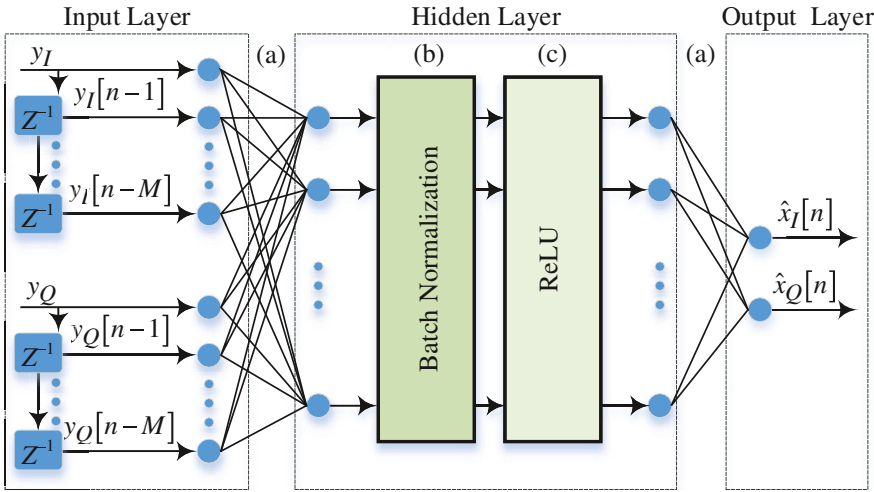
FIGURE 3 Structure of the neural network architecture proposed in this paper. The hidden layer consists of two dense layers, which are the batch normalisation layer and the ReLU layer

TABLE 1 Algorithm used for NN training

Algorithm 1 NN Training process
i: Generate n samples of IQ data of $x[n]$ and $y[n]$
ii: Update weights and biases given by Equation (4) using 70% of n samples
iii: Continue updating until finding the optimum $u(t) = h^{-1}(t)$
iv: Validate the model using 30% of n samples
v: If the cost function of validation is ok, then freeze the model
vi: Construct the pre-distorted signal, $x(t)*h^{-1}(t)$

Abbreviation: NN, neural network.

and ReLU are evaluated as activation functions in this work, where ReLU has been chosen as the activation function due to its performance, which is described in Section 4. ReLU introduces a non-linearity by deactivating negative inputs, adding sparsity to the model, and accelerating convergence [31]. ReLU is defined as:

$$f_{\text{ReLU}}(u) = \max(0, u), \tag{6}$$

where u is the input to the activation function. In this way, the output of a hidden layer can be expressed as Equation (7):

$$\mathbf{x}_{(i+1)} = \max(0, \hat{\mathbf{y}}_{(i)}), \tag{7}$$

where $\mathbf{x}_{(i+1)}$ is the input of the next FC layer. With the sequential structure, the inputs of the subsequent hidden layer can be described in terms of the current hidden layer. This procedure that goes from the first layer to the last layer is called forward propagation.

3.5 | Cost function

There are different ways, based on the type of the problem, to evaluate the difference between the real output and the estimated output, the so-called cost function. In this work, the

effects of two kinds of cost functions, Huber cost and mean-square-error (MSE) cost have been investigated. In the Huber cost function, instead of minimising the cost function, $|x_i - \hat{x}_i|$, the smooth cost function, L_1, is used for regression because it is robust against gross errors [32]. The smooth L_1 cost function is defined as

$$L_1(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{B} \sum_{i=1}^{B} \varepsilon_i, \tag{8}$$

where B is the batch size and ε_i is defined as a combination of the squared error and absolute error

$$\varepsilon_i = \begin{cases} 0.5(x_i - \hat{x}_i)^2, & \text{if} \quad |x_i - \hat{x}_i| < 1, \\ |x_i - \hat{x}_i| - 0.5, & \text{otherwise.} \end{cases} \tag{9}$$

The corresponding MSE cost function is defined as:

$$\text{MSE}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{B} \sum_{i=1}^{B} (x_i - \hat{x}_i)^2. \tag{10}$$

4 | NN TRAINING AND PARAMETER TUNING

4.1 | Training process

The concept of the proposed application is illustrated in Figure 1 and the configuration of the designed NN is shown in Figure 3. As shown in Figure 1, the parameters of the NN are updated step by step by reducing the losses between outputs of the NN (i.e., predicted values) and the reference inputs. The NN can gradually learn features hidden in training data for classification or regression missions. Generally, if the NN is trained as a classifier, the cross-entropy function is a commonly used cost function. For linearisation of active circuits where the NN needs to learn

the distinctive non-linear distortions, it is trained as the regression model.

Considering the memory effect of the active array, the memory depth, M, has a direct impact on the number of neurons in the input layer. So, there is a trade-off between the size of M and the linearisation performance. Since the output training data and the reference input data are complex values (I and Q), the number of neurons of the input layer and output layer is set to 2 M and 2, respectively.

Adjacent channel leakage ratio (ACLR) and error vector magnitude (EVM) are used as metrics for choosing the desired parameter values in each training step. The ACLR describes the power of the leakage in the adjacent channel compared to the in-band channel power and is defined as:

$$\text{ACLR}_{\text{dB}} = 10\log_{10}\left(\frac{P_{\text{adj}}}{P_{\text{in-band}}}\right), \tag{11}$$

where P_{adj} and $P_{\text{in-band}}$ are the powers of the adjacent channel and the main channel, respectively. In this way, the signal integrity can be directly assessed in the frequency domain. The left side of ACLR is used for evaluation through the experiments in this paper. Since ACLR only measures the distributed power in different channels, another metric for in-band signal quality, EVM, in terms of percentage, is calculated as:

$$\text{EVM}_{\%} = \sqrt{\frac{P(\text{in - band})_{\text{error}}}{P(\text{in - band})_{\text{ref}}}} \cdot 100_{\%}, \tag{12}$$

where $P(\text{in-band})_{\text{error}}$ and $P(\text{in-band})_{\text{ref}}$ are the powers of the error vector and ideal signal vector in I and Q planes, respectively.

All operations are realised using Python 3.8.4 on Visual Studio Code. The NN is built and trained using Keras 2.3.0-tf, and the version of Tensorflow is 2.2.0.

4.2 | Parameter tuning

For parameter tuning, 100 k I and Q samples of input and output of the active array are captured, where 70% of the data was randomly chosen for training and the remaining 30% for testing.

The memory depth, the number of neurons in each layer and the number of hidden layers should be set appropriately. If these numbers are too small, then NN cannot get the right features for the non-linearity model and there is a risk of underfitting, the same is valid if these numbers are too large, which results in overfitting. To avoid this, the memory depth has been chosen to a low number, and then, the other parameters have been initialised to get the best linearisation parameter in terms of ACLR and EVM. For achieving faster tuning and for reducing the number of

multiplications in a real application, 1 hidden layer is chosen. The optimisation was continued by tuning the memory depth and keeping other parameters unchanged. Table 2 shows how the NN is configured together with the ACLR and EVM results for the different number of memory depths. The best performance is achieved by setting the memory depth to five which results in the best ACLR improvement.

Having memory depth fixed to five, the other optimisation parameters such as activation function, cost function, batch size and the number of epochs have been tuned. The results are listed in Table 3 and based on those parameters, the ReLU and the MSE are chosen for the activation function and the cost function, respectively. Further evaluation on the number of epochs shows that approximately 25 epochs are enough for the algorithm to reach its minimum cost of 1E −6 and the cost function will not improve further as shown in Figure 4.

4.3 | NN simulation results

Figure 5 shows the simulation results of power spectral density (PSD), amplitude-to-amplitude (AM-AM) and amplitude-to-phase (AM-PM) distortions for the active array output. The parameters used for simulation are based on the best tuning parameters from Table 3.

Several sets of pre-distorted signals have been trained based on the final model and have been used for characterising the efficiency of the model versus the level of non-linearity in the active array. These results are discussed in the next section.

TABLE 2 Configurations and training for optimising memory depth

Optimiser	Adam				
Activation function	ReLU				
Cost function	Huber				
Accelerator	BN				
Hidden layers	1				
Neurons	100				
Epoch	50				
Batch size	100				
Initial learning rate	0.1				
Minimum cost	$1E - 6$				
Training data	70%				
Validation data	30%				
Memory depth	3	4	**5**	6	7
ACLR improvement	6.2 dB	6.3 dB	**6.5 dB**	6.2 dB	5.5 dB
EVM improvement	7.4%	8%	**8%**	8.1%	8%

Abbreviations: ACLR, adjacent channel leakage ratio; EVM, error vector magnitude.

5 | COMPLEXITY ANALYSIS

5.1 | MP pre-distortion

Equation (2) models the behaviour of the PA, which means that the APA output can be estimated from the inputs. For pre-distortion, the inverse model is needed, which means that the input should be estimated based on the output that is implemented by switching the input and outputs:

$$x[n] = \sum_{m=0}^{M} \sum_{k=1}^{K} a_{mk} y[n-m] \cdot |y[n-m]|^{k-1}, \quad (13)$$

where $x[n]$ is the estimated input to the APA. This can be written as a vector-vector product in the form:

TABLE 3 NN parameter optimisation. Parameters include activation function, cost function, batch size and epochs size

Memory depth = 5				
Parameter optimisation[a]				
Activation	Cost	Batch size	Epochs	ACLR improvement
ReLU	MSE	100	50	7.03 dB
ReLU	Huber	100	50	6.50 dB
ReLU	MSE	20	50	6.04 dB
Tanh	Huber	100	50	5.52 dB
ReLU	MSE	1000	500	3.03 dB
ReLU	MSE	1000	50	1.84 dB

Abbreviations: ACLR, adjacent channel leakage ratio; MSE, mean-square error; NN, neural network.
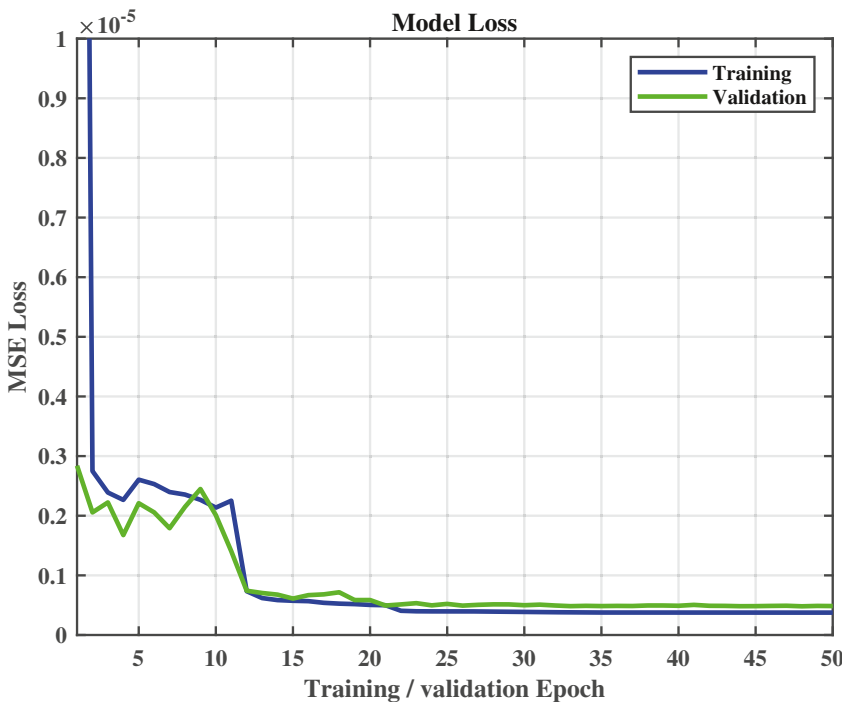
[a]Sorted based on decreasing ACLR improvement.

$$\mathbf{x} = \mathbf{rw}, \quad (14)$$

where

$$\mathbf{r} = \left[y[n-0] \cdot |y[n-0]|^{1-1} \quad \cdots \quad y[n-M] \cdot |y[n-M]|^{K-1} \right] \quad (15)$$

Since the absolute value, $|\cdot|$, requires three multiplications, the complexity of finding $\mathbf{r}$ is thus:

$$C_{r,complex} = (M+1) \cdot (0 + 3 + 4 + 5 + \cdots + k + 1)$$
$$= \sum_{i=3}^{K+1} i \cdot (M+1) \quad (16)$$

The complexity of the vector-vector product $x[n] = rw$ is simply:

$$C_{Mul,MP,complex,vector} = (M+1)K, \quad (17)$$

$$C_{Add,MP,complex,vector} = (M+1)K - 1. \quad (18)$$

where $C_{Mul,MP,complex, vector}$ is the number of complex multiplications and $C_{Add,MP,complex, vector}$ is the number of complex additions. The total number of complex multiplications becomes:

$$C_{Mul,MP,complex, total} = (M+1)K + \sum_{i=3}^{K+1} i \cdot (M+1). \quad (19)$$

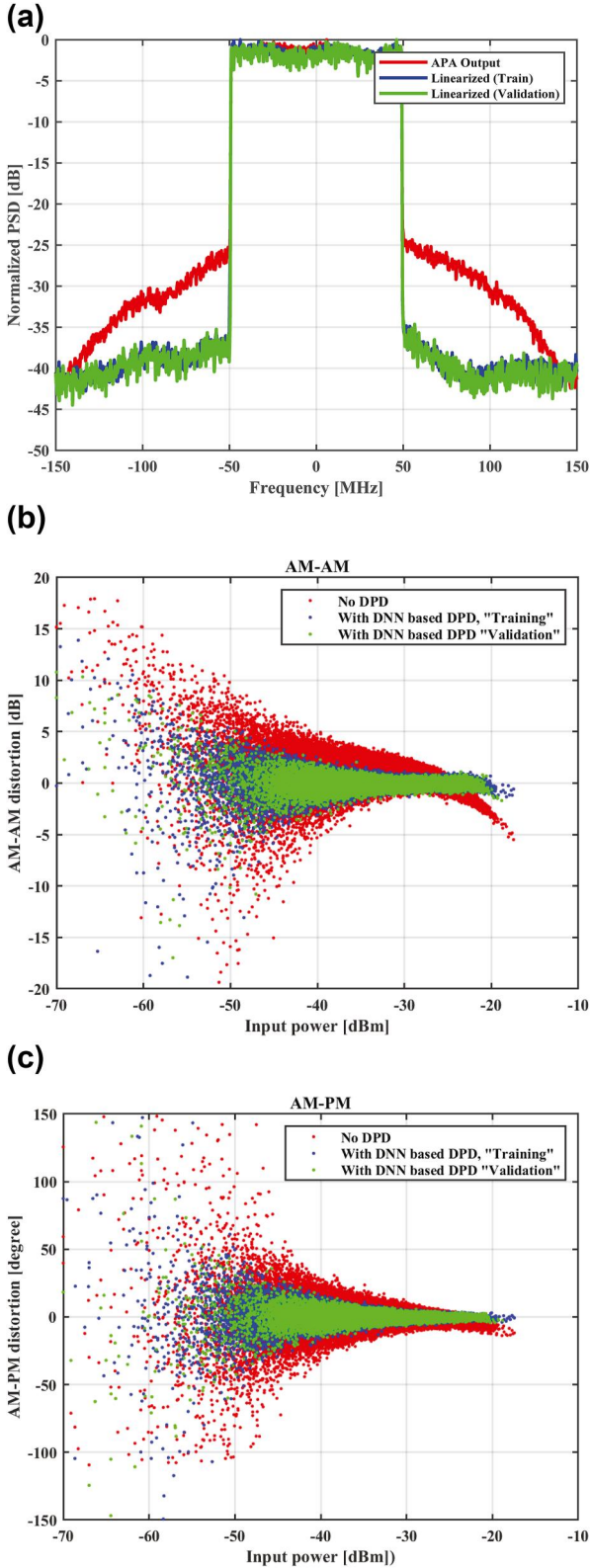**FIGURE 4** Model loss versus the number of training/validation epochs

FIGURE 5 Simulated linearisation results with and without NN pre-distortion: (a) Power spectral density; (b) amplitude-to-amplitude (AM-AM) distortion; (c) amplitude-to-phase (AM-PM) distortion

A complex multiplication takes four real multiplications and two real additions and a complex addition involves two real additions. This means that the total complexity of the MPM pre-distortion in real operations is:

$$C_{\text{Mul,MP}} = 4 \cdot \left[(M+1)K + \sum_{i=3}^{K+1} i \cdot (M+1) \right] \quad (20)$$

$$C_{\text{Add,MP}} = 2 \cdot \left(\left[(M+1)K + \sum_{i=3}^{K+1} i \cdot (M+1) \right] + \left[(M+1)K - 1 \right] \right) \quad (21)$$

5.2 | NN pre-distortion

The complexity analysis is made with a starting point in Equation (4) with L as the number of outputs of the previous layer, and P as the number of inputs to the next layer. If only fully connected layers with equal amounts of neurons are considered, the problem can be further reduced as $P = L$. Between each fully connected layer, there are P^2 multiplications and P^2 additions. The number of operations between the input layer and the first hidden layer is $2MP$ multiplications and additions, where M is the memory depth. Between the last hidden layer and the output layer, there are $2P$ multiplications and additions. Thus, the total amount of multiplications and additions is:

$$C_{\text{Mul,NN}} = C_{\text{Add,NN}} = 2MP + (J-1)P^2 + 2P, \quad (22)$$

where J is the number of hidden layers. Equation (22) shows that complexity scales quadratically with the number of neurons if there is more than one hidden layer. The complexity grows linearly with the number of neurons if only a single hidden layer is used. According to the universal approximation theorem, a single hidden layer can be used for arbitrary function approximation, so for applications where low complexity is required, a single hidden layer may be desirable.

5.3 | Complexity comparison

For the MPM, the pre-distorted signals based on Equation (2) with various non-linearity orders, K, and memory depths, M, have been constructed and evaluated in the lab and the optimal values of $K = 5$ and $M = 8$ have been chosen. The NN is trained to reach the minimum MSE of approximately 1E-6 as an example shown in Figure 4. By sweeping the NN parameters, one hidden layer and 100 neurons and a memory depth of five have been chosen. Table 4 shows the computational effort in terms of multiplications and additions based on Equations (20)–(22). Although the number of multiplications is higher in the case of NN compare to MPM, the absolute number is still very low, and besides, NN has superior linearisation performance, which is shown in Section 6.

6 | MEASUREMENT RESULTS

6.1 | OTA measurement setup

The block diagram of the measurement setup is shown in Figure 6 and the actual laboratory measurement setup is illustrated in Figure 7.

The R&S SMBV100 B Vector Signal Generator and its arbitrary waveform generator function generate the TX input IF signal, centred at 3 GHz, which is a 100 MHz bandwidth 5G NR signal. It is a 3GPP downlink OFDM modulated waveform with 64-QAM sub-carrier modulation, sub-carrier spacing of 60 kHz and 1584 active sub-carriers. With an oversampling factor of 6, the sample rate of the transmitter and receiver signals is 600 MHz. The peak-to-average power ratio (PAPR) of the input signal, after capturing and loading to the generator, is 11.6 dB. For up-conversion and down-conversion, an un-modulated signal of 12.5 GHz has been generated by an Agilent E3247 C and frequency-doubled to 25 GHz using a MITEQ-MAX2M200400 and fed into a power divider to be used as a local oscillator (LO) signal. Two active mixers, KTX321840 and KRX321840, operating in their highly linear region, are utilised for up converting the IF signal to the 28-GHz carrier frequency and for down-converting the signal back to IF. A 28-GHz

band-pass filter is used to select the up-converted modulated signal and suppress the LO leakage and image frequency signals. The Ducommum APH-26 063 325 is used as a pre-amplifier. The pre-amplifier is a high-power device and while operating more than 10 dB below its compression point, the output is linear and the power is sufficient to drive the 4 × 4 APA, AAiPK428GC-A0404 [33], close to its saturated region. The APA includes four Anokiwave AWMF-0158 [34] and integrates 16 branches of attenuators, phase shifters, PAs and 16 patch antennas in a 4 × 4 active phased array. The APA is designed for a typical main beam power of +33 dBm.

The diagonal length of the active array antenna is approximately 4 cm, which at 28 GHz results in a far-field distance of:

$$\frac{2D^2}{\lambda} = 30.5 \quad \text{cm}, \tag{23}$$

where D is the diagonal length of the antenna and λ is the wavelength. The main beam signal is captured by the observation horn antenna placed 55 cm away, which is well above the far-field distance of the device.

After down-conversion to IF, the signal is captured by the R&S FSW Signal and Spectrum Analyser and converted to the base-band. A host PC running in Matlab and using the R&S ARB Toolbox is used for capturing and uploading the I and Q samples. The measurement setup is power calibrated in order to keep all other components in their linear operating regions and the only source of non-linearity is related to the active phased array. For controlling the main beam of the array, the code-book and software tools from Amotech [33] have been used. For MPM, 4 pre-distorted signals, one for each power level, have been constructed based on a memory depth of 8 and a non-linearity order

TABLE 4 Computational effort in terms of multiplications and additions for MPM and NN

	K	M	Layer	Neurons	Multiplications	Additions
MPM	5	8	-	-	828	502
NN	-	5	1	100	1200	1200

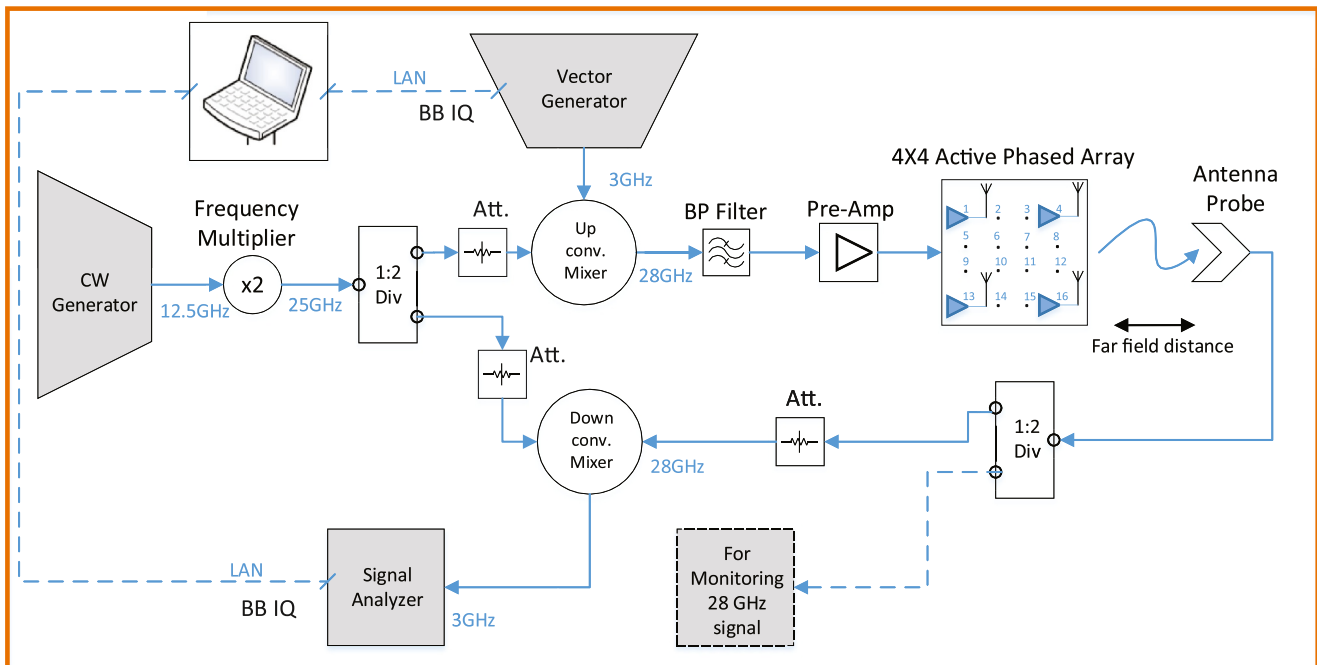Abbreviations: MPM, memory polynomial model; NN, neural network.

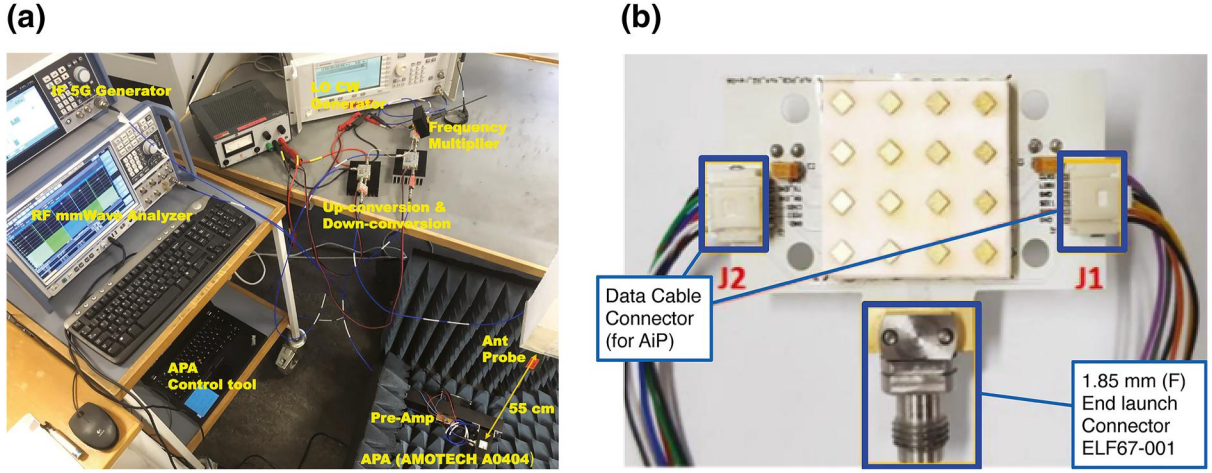**FIGURE 6** The block diagram of the measurement setup for the 4 × 4 array

FIGURE 7 (a): Measurement setup in Lab; (b): AAiPK428GC-A0404 evaluation board used for measurements [33]

of five. For NN, four corresponding pre-distorted signals based on the parameters in Table 3 have been constructed and used as the input to the APA for the measurements.

6.2 | OTA measurement results

In this section, we present the experimental results of using the NN-based DPD and compare that with the MPM-based DPD. In both cases, we do the measurements on four different power levels where the APA is driven into compression. The radiated power has been adjusted to get the ACLR slightly worse than the limit for systems operating at FR2, which is −28 dBc [35]. The main beam power of the APA is the sum of the transmitter power and gain of the antenna array:

$$P_{\text{main-beam}_{APA}} = P_{\text{TX}_{APA}} + G_{\text{Ant}_{APA}}, \quad (24)$$

and is measured as the power at the observation horn antenna placed 55 cm away, adding the propagation loss and subtracting the gain of the observation horn antenna.

The pre-distorted signals are fed as input to the APA and the corresponding ACLRs and EVMs are measured for each case and each technique.

The APA is driven in four different linearity cases with main beam power to be 34, 33, 32 and 31 dBm for very high, high-, medium- and low-power cases, respectively. Due to the limited output power of the 4 × 4 APA, we are not able to increase the main beam power further due to the risk of damage to the device. However, with the main beam power of 34 dBm, the device is in the saturated region and the ACLR is approximately 2 dB worse than the 3GPP limit and suitable for our analysis. Measured OTA spectra with NN and MPM-based linearisation for four different cases are shown in Figure 8 and the ACLR and EVM improvements are illustrated in Figure 9. Here, we can see that for higher non-linearities, the NN is performing better than MPM. In the case of low non-linearity, the MPM performance is equal or slightly better, but it is worth pointing out that linearisation is less meaningful for relatively linear and less power-efficient operations. The results

demonstrate that the proposed NN is capable of effectively learning the distinctive highly non-linear distortions, which may not easily fit to existing MPM solutions.

6.3 | ACLR and total radiated power (TRP)

Even though an existing reference claims the distortion is beam-formed in the same direction as the intended signal with a multi-antenna transmitter in the single-user case [4], quantitative results are desired to evaluate if using main beam Adjacent channel leakage ratio (ACLR) is a valid method for characterising the linearisation performance of a beam steerable array. For evaluating this, we performed a total radiated power (TRP) ACLR measurement and compared the results with main beam direction measurements. The TRP is defined from the integration of signal power over the angular domains. The estimated TRP for a discrete set of measured directions is defined as [35]:

$$\text{TRP}_{\text{Estimate}} = \frac{\pi}{2NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \text{EIRP}(\phi_n, \theta_m) \cdot \sin \theta_m, \quad (25)$$

where N and M are the number of azimuth angles, ϕ_n, and elevation angles, θ_n, respectively, and $\text{EIRP}(\phi_n, \theta_m)$ is the radiated power in each angular case as a sum of both linear polarisations. The TRP-ACLR in a linear scale is calculated as total radiated power of the adjacent channel divided by the total radiated power of the in-band channel:

$$\text{TRP - ACLR} = \frac{\frac{\pi}{2NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \text{P}_{adj.ch.}(\phi_n, \theta_m) \cdot \sin \theta_m}{\frac{\pi}{2NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \text{P}_{ch.}(\phi_n, \theta_m) \cdot \sin \theta_m}$$

$$= \frac{\sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \text{P}_{adj.ch.}(\phi_n, \theta_m) \cdot \sin \theta_m}{\sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \text{P}_{ch.}(\phi_n, \theta_m) \cdot \sin \theta_m}$$

$$(26)$$
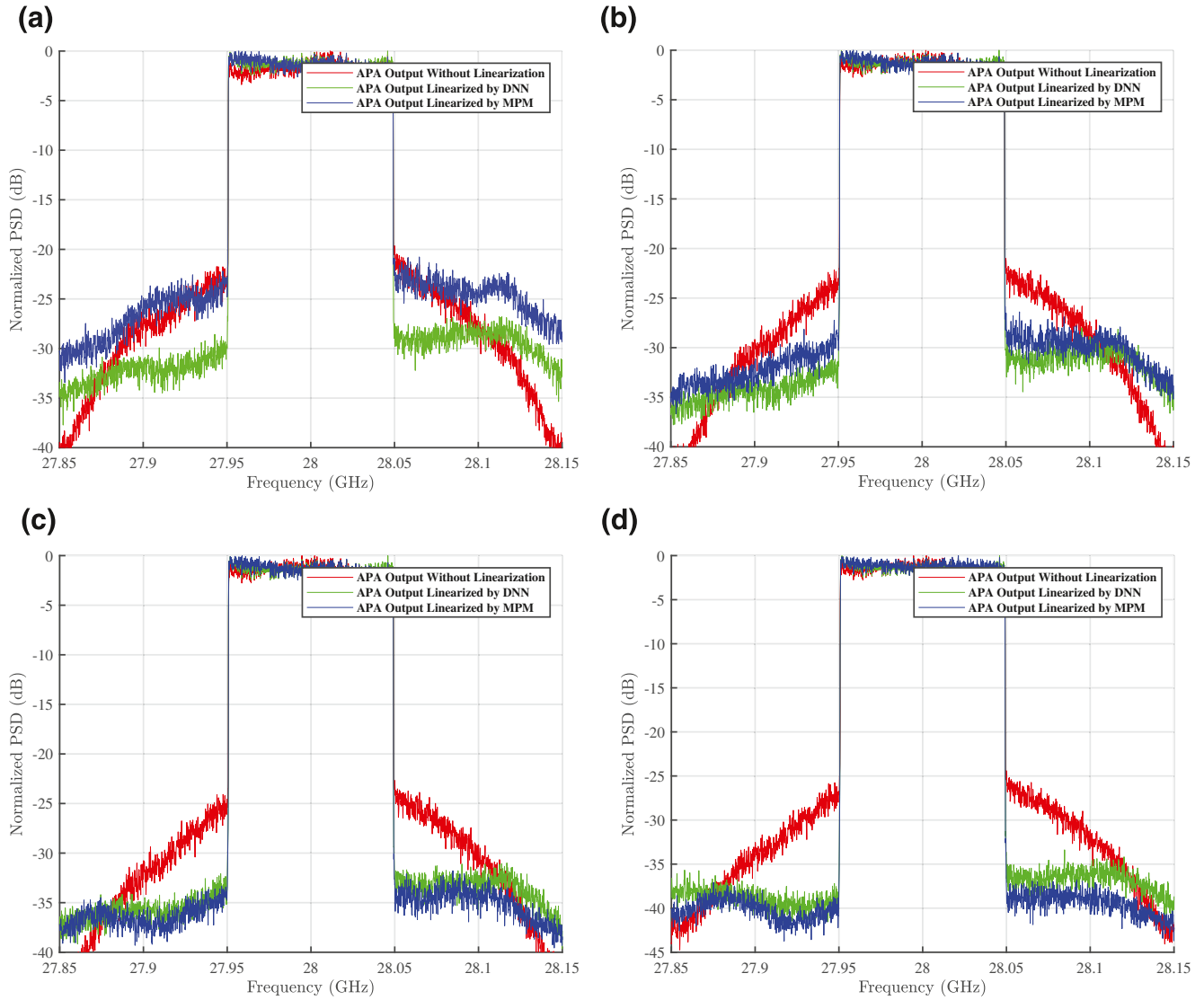
(a)

(b)

(c)

(d)

F I G U R E 8 Measured over-the-air spectra with NN and memory polynomial model (MPM)-based linearisation at different main beam power: (a) Very high non-linearity case (main beam power = +34 dBm); (b) High non-linearity case (main beam power = +33 dBm); (c) Medium non-linearity case (main beam power = +32 dBm); (d) Low non-linearity case (main beam power = +31 dBm). APA, active phased array; DNN, delay neural network; PSD, power spectral density

The block diagram and lab setup for the measurements are shown in Figures 10 and 11. The following procedure is applied for all specific angles θ and ϕ:

1. Place the APA at the positioner and align the coordinate system.
2. Align the beam of the APA to the desired beam steering angle.
3. Measure the main channel power and adjacent channel power using a spectrum analyser.
4. Repeat steps 1-3 for all directions in the TRP measurement grid.
5. Calculate TRP-ACLR according to Equation (26).

The steering angle of the APA is set to 0°. The position of the APA is changed by θ from 0 to 180° in steps of 10° and for each step, the ϕ angle is changed from −90 to 90° in 20 steps. The in-band channel power and the ACLR power for each angular position have been measured and the TRP-ACLR level has been calculated according to Equation (26) to be −35.0 dBc. For the same setup, the main-beam-only level of ACLR is measured to be −33.3 dBc. In our work, we assume the main beam pointing is maintained in communication. Moreover, since the measured difference is less than 2 dB, the main beam ACLR is chosen as the metric for experimental validation in this work.

6.4 | ACLR and beam directions

Due to interactions between the PAs in the array, the linearised beam is sensitive to the steering angle, so the impact of beam
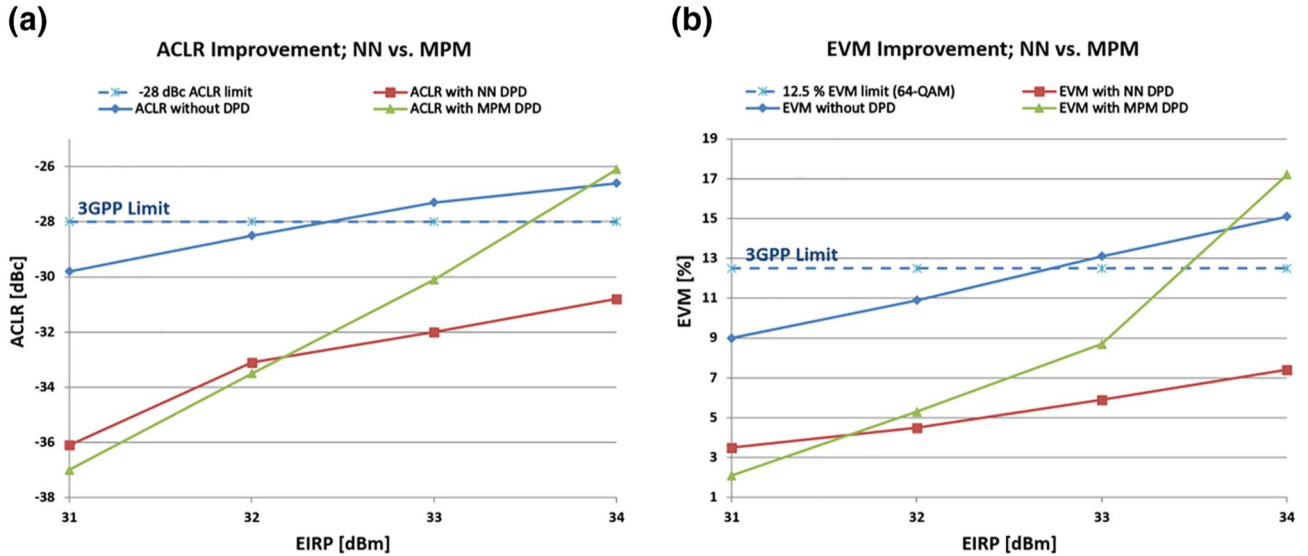
FIGURE 9 Comparison of neural network (NN) versus memory polynomial model (MPM): (a) adjacent channel leakage ratio (ACLR) comparison; (b) error vector magnitude (EVM) comparison
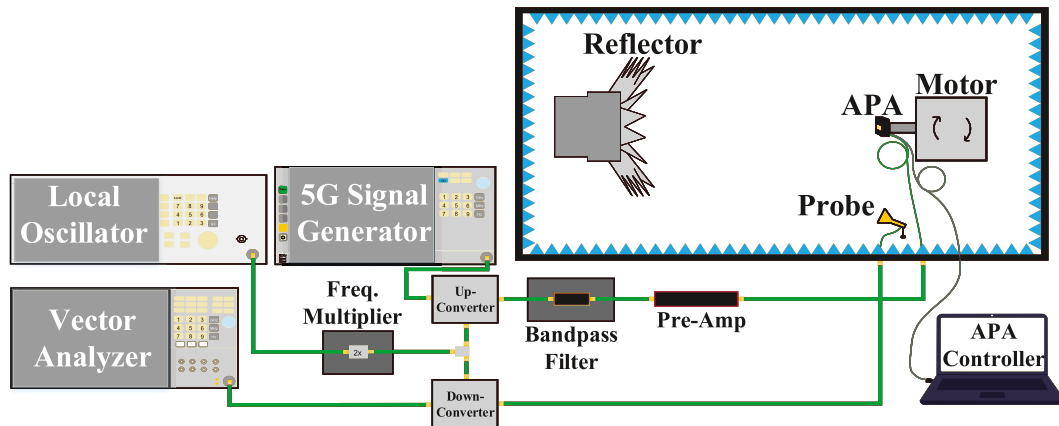

FIGURE 10 Block diagram of the measurement setup for measuring active phased array (APA) in a compact antenna test range

steering needs to be quantified. A single trained DPD is not sufficient for maintaining a low distortion in a wide range of steering angles. To maintain a low level of distortion across the steering angle, a new training after some degrees shift (based on the actual setup) of the main beam is required [8]. The same investigation has been done in [3] with the same conclusion. Furthermore, remarkable results in [36] show that the NN is capable of modelling the correlation between the non-linear distortion characteristics among different beams. This allows providing consistently good linearisation regardless of the beamforming direction, thus avoiding the necessity of executing continuous digital pre-distortion parameter learning. In this work, we have quantified the load modulation impact by measuring on an over-the-air test setup in a compact antenna test range (CATR). Figure 10 shows the setup used for measuring the APA output over the air in the CATR. In this experiment, 17 different values of the steering angle θ were used in the range of -78 to $+78°$ with a step of approximately $8°$. The following procedure is applied firstly to capture the

non-linear data for all angles and secondly the linearised data for all angles. For all specific angles θ_1 to θ_{17}, the following steps are used:

1. Adjust the steering angle to θ_i according to code-book and software tools.
2. Adjust the mechanical angle accordingly.
3. Measure input/output data for each steering angle.
4. Make the MPM and the NN pre-distorters based on pre-distortion coefficients obtained from measurements of the $0°$ steering angle.
5. Use the pre-distorter as input and repeat steps 1-3.

The results are shown in Figure 12, as measured for the same power level corresponding to the highly non-linear use case depicted in Figure 8b. The ACLR of the APA without linearisation is varying with the steering angle due to changes in radiation patterns and because of load modulation. Furthermore, the ACLR improvement rate for the linearised signals is
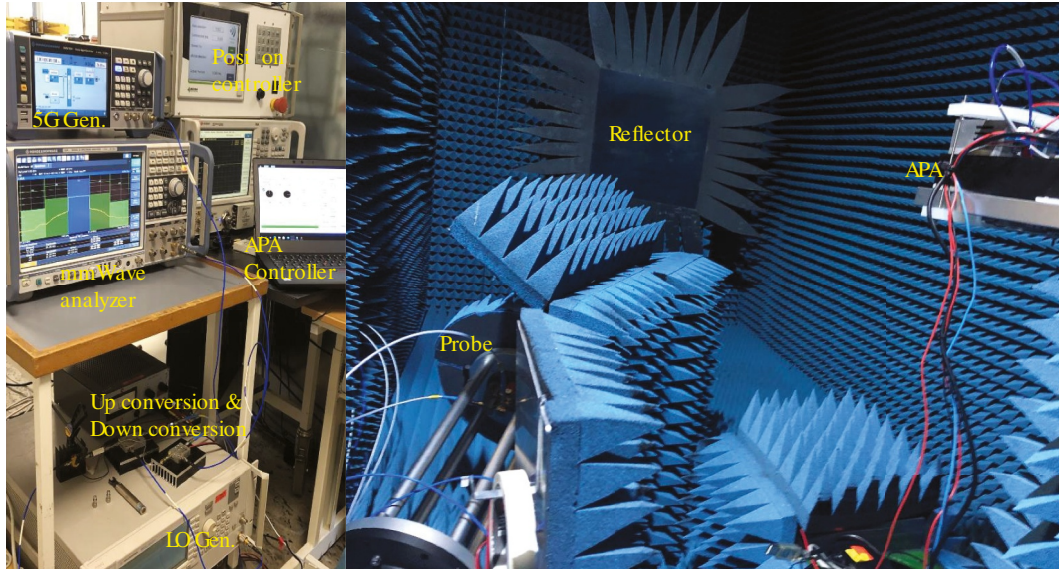
FIGURE 11 Total radiated power measurement setup using compact range chamber. APA, active phased array
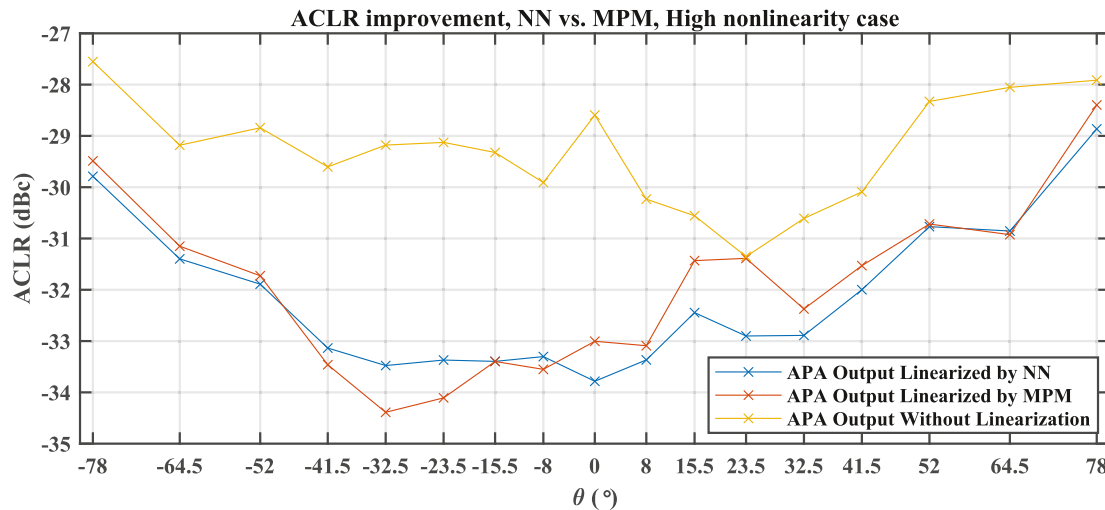

FIGURE 12 Measured adjacent channel leakage ratio (ACLR) performance of memory polynomial model (MPM) and neural network (NN)-based digital pre-distortion versus steering angle using the over-the-air setup in a compact range chamber

varying with the steering angle also as a result of load modulation.

6.5 | Time-domain comparison of NN and MPM pre-distortion signals

In this section, we compare the pre-distorted signals of the NN and MPM DPD to understand better what the NN does differently. A possible explanation for this can be found by inspecting the signals in the time domain. Figure 13 shows the complex envelope of the reference input signal, the non-linear signal, the pre-distorted signal and the response after pre-distortion for the two power levels, 31 and 34 dBm indicated as low and high linearity cases, respectively. The gain is normalised to 0 dB for comparison.

The pre-distorted signal, as expected, has extra gain to counteract the decreasing non-linear gain at the points where the non-linear signal is in compression, which is illustrated as the high peaks in the time domain. As a consequence when the pre-distorted signal is applied, the response of the pre-distorted signal should ideally end up on top of the reference signal. This is exactly what happens in Figure 13a,b, where there is almost no difference between the reference input signal and the measured response after pre-distortion. The case of high non-linearities can be seen in Figure 13c,d, where the difference between the reference input signal and the response after pre-distortion can now be easily observed. It is clearly seen from Fig. 13c that the MPM technique overcompensates the compression. When comparing the NN approach with the polynomial one, the polynomials have inherent local approximating properties in contrast to the global approximation
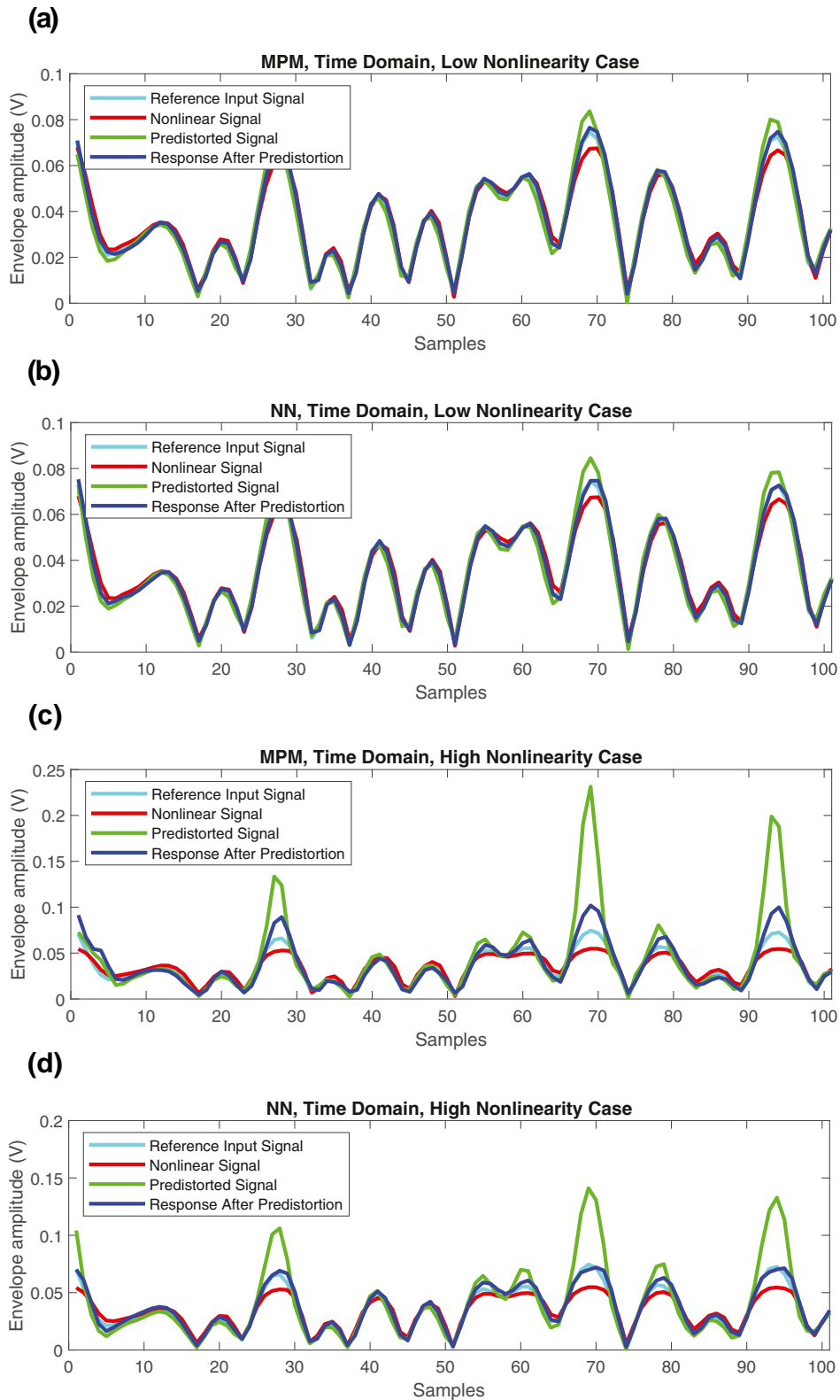
(a)

(b)

(c)

(d)

FIGURE 13 Time-domain representation of pre-distorted signal: (a) memory polynomial model (MPM) low non-linearity case; (b) neural network (NN) low non-linearity case; (c) MPM high non-linearity case; (b) NN high non-linearity case

capability of NNs when modelling strongly non-linear systems. Therefore, NN may adapt better to extrapolating beyond the zone exploited for parameter extraction [37].

Although this effect is still under the investigation, we see clearly the impact of the pre-distorted signal's PAPR on overall linearisation and, as a consequence, the shortcoming of the

conventional and less complex MPM for linearising highly non-linear 5G modulated signals, whereas a less complex NN-based technique can do the job satisfactorily.

7 | DISCUSSION

The phenomenon of differences in the PAPR in MPM versus NN DPD was observed for all measurements where the output power of the APA in our setup is above +32 dBm. During experiments, we kept the PAPR of the input signal as defined in 3GPP, that is, 11.3 dB without providing any clipping and filtering method to reduce the PAPR. The root cause of the difference between MPM and NN approaches can be explained as:

1. The MPM non-linearity kernel is a polynomial. For high non-linearities high-order polynomials are necessary. In our MPM, we used a non-linearity order equal to five and a memory depth of eight. When trying to linearise a highly compressed, deeply saturated PA characteristic, such a high-order polynomial non-linear function, quickly explodes at the upper side of the input amplitude range, thus causing the peaks of the pre-distorted signal to reach extremely high values, and hence leading to a huge increase in the pre-distorted signal PAPR compared to the PAPR of the original input-modulated signal.

2. The non-linear kernel of the NN does not contain inherently such an 'explosion' effect for high amplitudes. It is important to keep in mind that the proposed RVTDNN structure is based on supervised learning. While we train the NN, we are actually using low envelope fluctuations, that is, the desired I and Q at the output layer in Figure 3, which allows the NN to learn the characteristics for a signal with low envelope fluctuations. This is also related to the fact that the long-term memory is built into the RVTDNN through supervised learning. This kind of long-term memory can be used to simulate the slow dynamic changes of non-linear characteristics of the PA over time, mentioned in [21].

8 | CONCLUSION

This paper presents how a neural network (NN)-based linearisation technique behaves on the digital pre-distortion (DPD) of a highly non-linear active phased array (APA) using a wideband 3GPP 5G mmWave base-station transmitter signal and compares it to the used memory polynomial model (MPM) technique. The proposed design is implemented in a state-of-the-art 4×4 APA and a setup using up- and down-conversion from sub 6 to 28 GHz and having high non-linearity of the active phased array as the main impairment factor. The NN is built and trained using a Python simulation environment. The performance of the optimal NN pre-distorter was assessed with measurement results and compared to the MPM-based DPD technique. Measurement

results on the proposed NN technique show that in the case of very high non-linearity with an adjacent channel leakage ratio (ACLR) of −26 dBc, the pre-distortion signal generated by the NN exhibits peak-to-average power ratio (PAPR) much lower than the one generated by MPM and consequently is still capable to linearise the APA where it is not possible for the actual MPM technique. The proposed NN-based DPD technique applied on a highly non-linear APA with an ACLR of −28 dBc shows an improvement of error vector magnitude (EVM) of 7.2% points and ACLR of 4.7 dB. For the same setup, an MPM-based DPD can only achieve an improvement of EVM and ACLR of 4.4% points and 2.8 dB, respectively. In the future, we may include an investigation of the robustness of NN-based linearisation due to the steering angle and the impact of channel properties for the high non-linearity cases.

CONFLICT OF INTEREST
There isn't any conflict of interest.

DATA AVAILABILITY STATEMENT
The data used to support the findings of this study can be obtained from the Ph.D. student, Feridoon Jalili, at e-mail: fja@es.aau.dk.

ORCID
Feridoon Jalili https://orcid.org/0000-0002-3645-3728

REFERENCES

1. Gustavsson, U., et al.: Implementation challenges and opportunities in beyond-5G and 6G communication. IEEE J. Microw. 1(1), 86–100 (2021)
2. Alizadeh, M., Händel, P., Rönnow, D.: Behavioral modeling and digital pre-distortion techniques for RF PAs in a 3×3 MIMO system. Int. J. Microw. Wireless Technol. 11(10), 989–999 (2019)
3. Ng, E., et al.: Digital predistortion of millimeter-wave RF beamforming arrays using low number of steering angle-dependent coefficient sets. IEEE Trans. Microw. Theor. Tech. 67(11), 4479–4492 (2019)
4. Fager, C., et al.: Linearity and efficiency in 5G transmitters: New techniques for analyzing efficiency, linearity, and linearization in a 5G active antenna transmitter context. IEEE Microw. Mag. 20(5), 35–49 (2019)
5. Tervo, N., et al.: Digital predistortion of amplitude varying phased array utilising over-the-air combining. In: 2017 IEEE MTT-S International Microwave Symposium (IMS), pp. 1165–1168. IEEE (2017)
6. Liu, X., et al.: Beam-oriented digital predistortion for 5G massive mimo hybrid beamforming transmitters. IEEE Trans. Microw. Theor. Tech. 66(7), 3419–3432 (2018)
7. Jalili, F., et al.: Linearization of active transmitter arrays in presence of antenna crosstalk for 5G systems. In: 2019 IEEE Nordic Circuits and Systems Conference (NORCAS): NORCHIP and International Symposium of System-on-Chip (SoC), pp. 1–5. IEEE (2019)
8. Jalili, F., et al.: Linearization trade-offs in a 5G mmwave active phased array OTA setup. IEEE Access. 8, 110669–110677 (2020)

9. Hausmair, K., et al.: Modeling and linearization of multi-antenna transmitters using over-the-air measurements. In: 2018 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1–4. IEEE (2018)

10. Brihuega, A., et al.: Piecewise digital predistortion for mmwave active antenna arrays: algorithms and measurements. IEEE Trans. Microw. Theor. Tech. 68(9), 4000–4017 (2020)

11. Yu, C., et al.: Full-angle digital predistortion of 5G millimeter-wave massive mimo transmitters. IEEE Trans. Microw. Theor. Tech. 67(7), 2847–2860 (2019)

12. Brihuega, A., et al.: Mixture of experts approach for piecewise modeling and linearization of rf power amplifiers. IEEE Trans. Microw. Theor. Tech. (2021)

13. Ku, Y.: The volterra & wiener theories of nonlinear systems: Martin schetzen. 531 pages, diagrams, illustr., John Wiley, New York, 1980. Pergamon, New York (1982)

14. Eun, C., Powers, E.J.: A new volterra predistorter based on the indirect learning architecture. IEEE Trans. Signal Process. 45(1), 223–227 (1997)

15. Hornik, K.: Approximation capabilities of multilayer feedforward networks. Neural Network. 4(2), 251–257 (1991)

16. Wang, D., et al.: Augmented real-valued time-delay neural network for compensation of distortions and impairments in wireless transmitters. IEEE Trans. Neural Networks Learn. Syst. 30(1), 242–254 (2018)

17. Jaraut, P., Rawat, M., Ghannouchi, F.M.: Composite neural network digital predistortion model for joint mitigation of crosstalk, I/Q imbalance, nonlinearity in MIMO transmitters. IEEE Trans. Microw. Theor. Tech. 66(11), 5011–5020 (2018)

18. Zhang, Y., et al.: Vector decomposition based time-delay neural network behavioral model for digital predistortion of RF power amplifiers. IEEE Access. 7, 91559–91568 (2019)

19. Vuolevi, J.H., Rahkonen, T., Manninen, J.P.: Measurement technique for characterizing memory effects in RF power amplifiers. IEEE Trans. Microw. Theor. Tech. 49(8), 1383–1389 (2001)

20. Chen, Q., et al.: Robust Digital Signal Recovery for LEO Satellite Communications Subject to High SNR Variation and Transmitter Memory Effects. IEEE Access (2021)

21. Liu, T., Boumaiza, S., Ghannouchi, F.M.: Dynamic behavioral modeling of 3g power amplifiers using real-valued time-delay neural networks. IEEE Trans. Microw. Theor. Tech. 52(3), 1025–1033 (2004)

22. Rawat, M., Rawat, K., Ghannouchi, F.M.: Adaptive digital predistortion of wireless power amplifiers/transmitters using dynamic real-valued focused time-delay line neural networks. IEEE Trans. Microw. Theor. Tech. 58(1), 95–104 (2009)

23. Liu, C., et al.: Algorithms for Verifying Deep Neural Networks. arXiv preprint arXiv:190306758, (2019)

24. Ghannouchi, F.M., Hammi, O., Helaoui, M.: Behavioral Modeling and Predistortion of Wideband Wireless Transmitters. John Wiley & Sons (2015)

25. Morgan, D.R., et al.: A generalized memory polynomial model for digital predistortion of RF power amplifiers. IEEE Trans. Signal Process. 54(10), 3852–3860 (2006)

26. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. 9(8), 1735–1780 (1997)

27. Gilabert, P.L., et al.: Machine learning for digital front-end: a comprehensive overview, pp. 327–381. Machine Learning for Future Wireless Communications (2020)

28. Cybenko, G.: Approximation by superpositions of a sigmoidal function. Mathematics of control, signals and systems. 2(4), 303–314 (1989)

29. Kingma, D.P., Ba, J.: Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:14126980, (2014)

30. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning, pp. 448–456. PMLR (2015)

31. Glorot, X., Bordes, A., Bengio, Y.: Deep sparse rectifier neural networks. In: Proceedings of the fourteenth international conference on artificial intelligence and statistics, pp. 315–323. JMLR Workshop and Conference Proceedings (2011)

32. Huber, P.J.: Robust estimation of a location parameter. In: Breakthroughs in statistics, pp. 492–518. Springer (1992)

33. Amotech. AAiP28G-A0808-EVB-R01_190321c (2020)

34. Anokiwave AWMF-0158 28 GHz silicon 5G Tx/Rx quad core IC. (2019) https://www.anokiwave.com/products/awmf-0158/index.html

35. Access, E.U.T.R.: Base station radio transmission and reception, vol. 36, pp. 96–97 (2018)

36. Brihuega, A., Anttila, L., Valkama, M.: Neural-network-based digital predistortion for active antenna arrays under load modulation. IEEE Microw. Wireless Compon. Lett. 30(8), 843–846 (2020)

37. Gilabert Pinal, P.L.: Multi Look-Up Table Digital Predistortion for RF Power Amplifier Linearization. Universitat Politecnica de Catalunya (2008)