# Review of the Neural Network based Digital Predistortion Linearization of Multi-Band/MIMO Transmitters

Praveen Jaraut[1], Mohamed Helaoui[1], Wenhua Chen[2], Meenakshi Rawat[3], Noureddine Boulejfen[4], Fadhel M. Ghannouchi[1]

[1] Intelligent RF Radio Laboratory, Schulich School of Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada
[2] Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
[3] Department of Electronic and Communication Engineering, Indian Institute of Technology, Uttarakhand, 247667, India
[4] Military Research Center, Ministry of Defense, 2045, Tunis, Tunisia

2021 IEEE MTT-S International Wireless Symposium (IWS) | 978-1-6654-3527-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/IWS52775.2021.9499466

***Abstract*—This paper reviews the neural network (NN) based digital pre-distortion techniques for linearizing multi-band and multiple-input multiple-output (MIMO) transmitters. The most popular NN-based behavioral models using Deep Neural Networks (DNN), Augmented Neural Networks (ANN), and Shallow Neural Networks (SNN) will be reviewed. The paper will discuss NN-based DPD models using Real-Valued Focused Time-Delay Neural Network (RVFTDNN) and Convolutional Neural Network (CNN) performance and complexity for multi-band and MIMO applications. These models' performance will be assessed in terms of their capability to mitigate the transmitter's distortion and hardware impairments such as crosstalk, PA's nonlinearity, dc offset, and *I/Q* imbalance for multi-band and MIMO applications.**

***Index Terms*—5G mobile communication, crosstalk, digital predistortion, linearization, multi-band transmitter, MIMO transmitter, power amplifier.**

## I. Introduction

In fifth-generation (5G) communication, signals having high data rates with large bandwidth are required in deployment scenarios such as those in dense urban, urban macro, and rural areas [1]. The Transceivers should support multi-standards, multiple bands, and multiple-input multiple-output (MIMO) topologies to achieve a high data rate using 4G Long-Term Evolution advanced (LTE-A) and 5G-New-Radio (NR) 5G signals. These signals suffer from high peak-to-average power ratio (PAPR). Power Amplifiers (PAs) in downlink transmitters of base stations (BSs) deliver high power with relatively broad signal bandwidth. PAs should work as close as possible to the compression region to achieve high efficiency [2]. However, in such conditions, they behave non-linearly and generate in-band and out-of-band distortion. These nonlinearities further increase in signals with high PAPR and large bandwidth, particularly in MIMO and multi-band applications [3].

Nowadays, digital predistortion (DPD) is a widely used technique for PA's linearization in the downlink transmission. This is mainly due to its relatively excellent performances in SISO transmitters as well as in multi-band and MIMO software defined radios (SDRs) [4]-[9]. Neural Network (NN) based models improve the modeling performance in terms of the nonlinear prediction of a PA's output signal, as compared to polynomial based models [4]-[9]. NNs are divided into shallow (SNN) and deep neural networks (DNN). SNN based DPDs like real-valued recurrent neural network (RVRNN) [4], real-valued focused time-delay neural network (RVFTDNN) [5]-[6], and augmented convolutional neural real-valued time-delay neural network (ARVTDNN) [7] have one or two hidden layers with a small numbers of neurons in each layer. In [5], it is shown that RVFTDNN has better linearization performance as compared to RVRNN. RVFTDNN DPD is used to mitigate dc offset, *I/Q* imbalance, PA nonlinearity and crosstalk imperfections simultaneously in single band and concurrent dual-band transmitters [5]-[6]. ARVTDNN DPD has less complexity compared to RVFTDNN DPD [7]. The Augmented Convolutional Neural Network (ACNN) based DPD is proposed for multi-band transmitters using faster convolutional neural network architecture with less complexity [8]. In [9], NN based DPD is presented as a primary solution to compensate for dc offset, *I/Q* imbalance, PA nonlinearity and crosstalk imperfections simultaneously in MIMO transmitters.

The paper is organized as follows: Section II describes the NN based DPD model for mitigation of imperfections in multi-band transmitters. Section III describes the NN based DPD model for mitigation of imperfections in MIMO transmitters. Section IV gives the conclusion.

## II. NN based DPD Model for mitigation of imperfections in Multi-band transmitters

RVFTDNN and ARVTDNN are shallow NN based behavior models. Both models have real-valued time delay input signals and can be used to linearize multi-band transmitters. Fig. 1(a) shows the architecture of RVFTDNN model for multi-band transmitters. In this model, the input of the network consists of present and past values of the real baseband inputs $I_1$, $I_2$,..., $I_S$, $Q_1$, $Q_2$, and $Q_S$ of the $S$-dimensional ($S$-D) multi-band signals [5]-[6], where $S$ is the number of bands or carriers. However the ARVTDNN model consists of present and past values of the real baseband inputs $I_1$, $I_2$,..., $I_S$, $Q_1$, $Q_2$, $Q_S$, and their corresponding magnitudes $|x_1|$, $|x_1|^2$, $|x_1|^3$,... $|x_S|$, $|x_S|^2$, and $|x_S|^3$ of the $S$-D multi-band signals [7]. Both RVFTDNN and ARVTDNN models follow the feedforward back propagation.

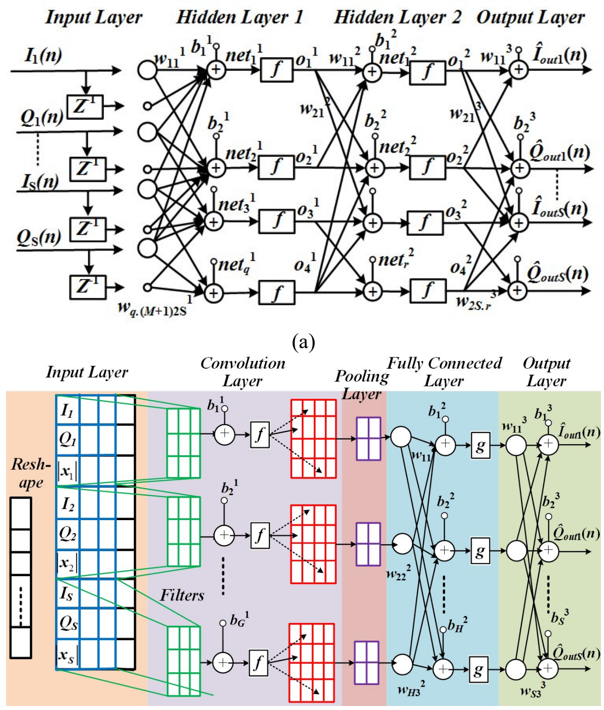


Fig. 1. Architecture of (a) RVFTDNN and (b) ACNN models for Multi-band transmitters.

## A. Feedforward Propagation

The data propagates from neurons of a lower layer to those of an upper layer in feedforward computation as shown in Fig. 1(a). The net input in layer $l$+1is given by

$$net_j^{l+1} = \sum_{i=1}^{q} w_{ji}^{l+1} o_i^{l} + b_j^{l+1} \tag{1}$$

where $b_j^{l+1}$ denotes the bias of the $j^{\text{th}}$ neuron in the $l$+1$^{\text{th}}$ layer. $q$ denotes the total number of neurons in the previous layer and $w_{ji}^{l+1}$ denotes the synaptic weight between the $i^{\text{th}}$ input from the previous layer to the $j^{\text{th}}$ neuron of the present layer. The output of neuron $j$ at the $l$+1$^{\text{th}}$ layer is

$$o_j^{l+1} = f\left(net_j^{l+1}\right) \tag{2}$$

The Hyperbolic tangent function is the activation function ($f$) of hidden layers. The outputs of the hidden neurons are linearly summed up at the output layer [9].

## B. Backward Propagation

The performance index $V$ for the NN in backward propagation is calculated as

$$V = \frac{1}{2}\sum_{n=1}^{N}\left\{e_n^T e_n\right\} \tag{3}$$

where $e$ is the error between the outputs from output-layer neurons of the NN model and the actual baseband outputs of the $S^{\text{th}}$ band of the transmitter [9].

The Levenberg-Marquardt algorithm [5] is used to minimize the parameter $V$ with respect to a parameter **u** which

TABLE I
DPD PERFORMANCES FOR CONCURRENT DUAL-BAND TRANSMITTER

| Models | LB | | UB | | Total No. of | |
|---|---|---|---|---|---|---|
| | NMSE (dB) | ACPR (dBc) | NMSE (dB) | ACPR (dBc) | Coeff. | FLOPs/sample |
| W/O DPD | -9.4 | -35.1 | -9.2 | -34.2 | N/A | N/A |
| 2D-MP DPD | -34.8 | -45.2 | -31.7 | -42.3 | 280 | 2238 |
| 2D-GMP DPD | -37.2 | -48.4 | -33.5 | -44.1 | 352 | 2814 |
| RVFTDNN | -38.4 | -48.8 | -34.8 | -45.3 | 727 | 1845 |
| ARVTDNN | -39.1 | -49.2 | -36.2 | -46.6 | 673 | 1629 |
| ACNN DPD | -39.6 | -49.7 | -36.9 | -47.3 | 187 | 1066 |

depends on the synaptic weights and biases. During the backward propagation, the weights and biases are updated to reduce the mean square error (MSE). The entire process is iterated till a good performance is accomplished by the NN.

The ARVTDNN model also uses the Levenberg-Marquardt optimization algorithm to train the NN [7]. The hyperbolic tangent (tanh) is adopted as an activation function in the hidden layers of the ARVTDNN model [7].

Fig. 1(b) shows the architecture of the multi-band ACNN model. The input of this network is $o^0$ which consists of present and past values of the real baseband $S$-D multi-band input signals $I_1$, $I_2$,..., $I_S$, $Q_1$, $Q_2$, $Q_S$ and their magnitude $|x_1|, |x_2|, \ldots$ and $|x_S|$ . This input signal is restructured into a two-dimensional matrix in the input layer. The input layer is followed by the convolutional layer. The 2D convolutional layer is the essential building block of a CNN that performs most of the computations [8]. As an example, let the dimension of the input layer be $A\times B\times 1$ and the dimensions of the convolutional filter size of the convolution layer be $F_1\times F_2$ while $G$ is the number of the convolution filters. Then the number of the filter's weights and biases of the convolution layer are $F_1\times F_2\times G$ and $G$ respectively. Fig. 1(b) shows the detailed operation of the convolutional layer. The input convolves to the filter's weights and then the bias is added to the convolution sum. This sum is the input of the activation function $f$. The final output of the convolution layer is $f\left(o^0 \otimes w^1 + b^1\right)$ , where $o^0$ are the inputs, $w^1$ are the filter's weights, $b^1$ are the biases and $f$ is the hyperbolic tangent function. The output of this function is one of the elements of the convolutional output matrix. This operation is repeated by sliding each filter with a stride along the width, and height of the input layer as shown in Fig. 1(b). The pooling layer follows the convolution layer. It is used to decrease the size of the convolutional output matrix. This further reduces the amount of computation required and the number of weights in the ACNN model. Each output of the average pooling layer drives all the neurons of the fully connected layer. The hyperbolic tangent function $g$ is used as the activation function of the fully connected layer. The final layer of the ACNN is the output layer. The output layer uses a linear function as the activation function. The CNN model uses the ADAM algorithm for the updating of the filter weights and backpropagation, leading to a fast processing [8].

Table I shows the DPD performance of different models in concurrent dual-band transmitters. ACNN DPD has better

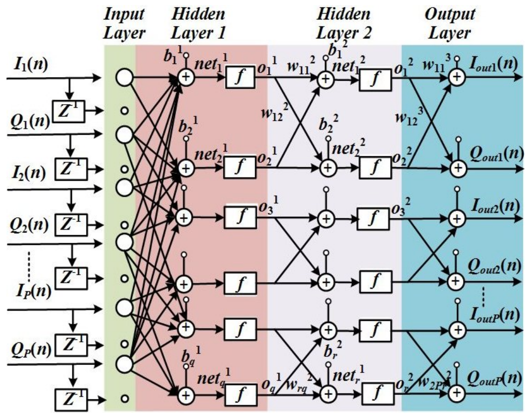


Fig. 2. Real-valued time-delay based Neural Network for MIMO Transmitter.

normalized mean square (NMSE) and adjacent channel power ratio (ACPR) as compared to other NN and polynomial based DPDs with less coefficients and floating point operations (FLOPs).

## III. NN based DPD Model for mitigation of imperfections in MIMO transmitters

### A. MIMO Transmitter

The gain imbalance, phase imbalance and LO leakage on the transmitted signal are the hardware impairments present in the Direct-conversion transmitters. When the signal is distorted due to multi-branch crosstalk and PA nonlinearity, the aforementioned impairments augments in MIMO transmitting at the same carrier frequency in different transmitters' paths [9].

### B. Crosstalk

The LO leakage and coupling effects between different transmitters' paths causes the crosstalk [9]. It can be characterized as nonlinear and linear. Nonlinear crosstalk happens before the PA while linear crosstalk takes place after the PA. The former crosstalk effect is amplified when the composite signal passes through a nonlinear PA. Generally, MIMO crosstalk has a coupling factor between -15 dB to -30 dB [9].

### C. I/Q Imbalance

The incongruity among the in-phase (*I*) and quadrature-phase (*Q*) signal paths in the modulator causes the *I/Q* imbalance in a MIMO transmitter [9]. Let $x_p(n)$ represent the baseband input signal in $p^{th}$ transmitter path of $P \times P$ MIMO transmitters, where $p=1, 2, \ldots, P$. Let $\theta_p$ and $\alpha_p$ represents the $p^{th}$ transmitter path's phase and gain imbalances respectively. $\theta_p = 0°$ and $\alpha_p = 1$ in a balanced modulator.

### D. NN based DPD Model for MIMO Transmitter

The architecture of real-valued time-delay based NN for MIMO transmitter is shown in Fig. 2 [9]. The input vector comprises past and present values of $I_1$, $I_2$,..., $I_P$, $Q_1$, $Q_2$, and $Q_P$ of the $P$ transmitter branches. The unit delay operator is represented by $z^{-1}$ in Fig. 2. The feedforward backpropagation neural network is used together with the Levenberg-Marquardt optimization algo-rithm to train the NN. The hyperbolic tangent (tanh) activa-tion function is also used in the hidden layers of the NN.

## IV. Conclusion

This paper discussed the architecture and algorithms used to train NNs in DPD. The NN based DPD architecture requires the current and delayed versions of the real valued baseband input signals. The RVFTDNN and ARVTDNN DPDs are able to mitigate imperfections such as dc offset, I/Q imbalance, PA nonlinearity and crosstalk in multi-band transmitters. Similarly, RVFTDNN based DPDs have less complex architecture while exhibit better linearization performances in MIMO transmitters. CNN based DPDs are promising architectures offering superior performances with less complexity in terms of number of coefficients, computing implementations and built-in adaptation to operating and environments and can be contemplate as Artificial Intelligence DPD based model. ACNN DPDs have better in-band and out-of-band linearization performances as compared to other NN and polynomial based DPDs in multi-band transmitters.

## References

[1] A. Maeder *et al.*, "A Scalable and Flexible Radio Access Network Architecture for Fifth Generation Mobile Networks," *IEEE Commun. Mag.*, vol. 54, no. 11, pp. 16–23, Nov. 2016.

[2] M. N. Patwary *et al.*, "The Potential Short- and Long-Term Disruptions and Transformative Impacts of 5G and Beyond Wireless Networks: Lessons Learnt From the Development of a 5G Testbed Environment," *IEEE Access*, vol. 8, pp. 11352-11379, Jan. 2020.

[3] V. Camarchia *et al.*, "The Doherty Power Amplifier: Review of Recent Solutions and Trends," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 2, pp. 559-571, Feb. 2015.

[4] D. Luongyinh and Y. Kwon, "Behavioral modeling of power amplifiers using fully recurrent neural networks," *IEEE MTT-S Int. Microw. Symp. Dig.*, pp. 1979–1982, Jun. 2005.

[5] M. Rawat, K. Rawat and F. M. Ghannouchi, "Adaptive Digital Predistortion of Wireless Power Amplifiers/Transmitters Using Dynamic Real-Valued Focused Time-Delay Line Neural Networks," *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 1, pp. 95-104, Jan. 2010.

[6] M. Rawat, K. Rawat, M. Younes and F. Ghannouchi, "Joint mitigation of nonlinearity and modulator imperfections in dual-band concurrent transmitter using neural networks," *Electron. Lett.*, vol. 49, no. 4, pp. 253-255, 14 Feb. 2013.

[7] 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.

[8] X. Hu, et al., "Convolutional Neural Network for Behavioral Modeling and Predistortion of Wideband Power Amplifiers," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, doi: 10.1109/TNNLS.2021.3054867.

[9] 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, Nov. 2018.