Reduced Complexity Real-Valued Time Delay Neural Network Based Behavioral Model of 5G RF Power Amplifiers

Reem AlNajjar
Department of Electrical Engineering
American University of Sharjah
Sharjah, United Arab Emirates
g00079249@aus.edu

Oualid Hammi
Department of Electrical Engineering
American University of Sharjah
Sharjah, United Arab Emirates
ohammi@aus.edu

***Abstract*—Neural networks appear as an increasingly attractive approach for the behavioral modeling and linearization of radio frequency power amplifiers. However, their relative complexity when compared to analytically defined models hinders their wide adoption in field deployed systems. This work tackles the complexity issue in neural networks by proposing a two-box structure using a look-up table followed by a real-valued time delay neural network (RVTDNN) in a Hammerstein-like structure in order to reduce the overall complexity of neural network based behavioral models without compromising their performance. The proposed model was evaluated using experimental data for a Doherty power amplifier driven by a 100MHz 5G test signal. The performance and complexity of the proposed model were compared to that of the standalone RVTDNN and the standalone augmented real-valued time-delay neural network (ARVTDNN). The results show that, compared to the standalone RVTDNN and ARVTDNN, the proposed model improves the NMSE especially for a reduced number of parameters. Hence, allowing for a better complexity and performance tradeoff.**

Keywords— 5G, behavioral modelling, distortions, memory effects, neural networks, power amplifier, look-up table

I. INTRODUCTION

The upcoming 6G technology, which promises exceptional speed, capacity, and reduced latency, is the next step in wireless communication growth. It is anticipated to outperform 5G by incorporating innovative features that will allow for new applications and services. PAs are essential components of wireless communication systems that increase the signal power prior to transmission. Reducing the impairments caused by power amplifiers (memory effects and nonlinear distortions) is necessary to achieve high-quality communication. Behavioral modeling is an effective method for predicting the output signal of the PA and assessing and analyzing its nonlinear behavior. Conventional PA models, including the Volterra model and its simpler versions, have been used to formulate mathematical relations between PAs input and output signals [1]. The high correlation between the polynomial basis functions in these models makes it difficult to improve the modeling performance, even when additional polynomial terms are included.

Neural networks based techniques have recently drawn a lot of interest in power amplifier behavioral modeling and predistortion because they are better than traditional polynomial-based techniques at capturing intricate nonlinear dynamics and memory effects [2] [3]. The real-valued time delay neural network (RVTDNN) is one of the architectures that effectively strikes a compromise between model complexity and performance [4]. More advanced neural network structures, including convolutional neural networks (CNNs) and bidirectional long short-term memory (BiLSTM) networks, have been proposed in the literature for a behavioral modeling and predistortion of nonlinear power amplifiers exhibiting memory effects [5]-[9]. These advanced models results in a significantly larger parameters count and frequently require significant computational resources and lengthy training times. Conversely, RVTDNN maintains a simpler structure with fewer parameters while still achieving competitive accuracy. Without requiring recurrent or convolutional layers, its time-delay structure of the input features vector efficiently simulates the temporal dependencies present in power amplifier behavior. This made RVTDNN models very useful for real-time applications and attractive for hardware deployment because of their lower complexity, greater convergence speed, and simpler construction.

An improved version of the traditional RVTDNN architecture is the augmented real-valued time delay neural network (ARVTDNN) which is intended to offer higher modeling accuracy without a significant computational overhead [10]. The ARVTDNN outperforms the simple RVTDNN in capturing higher-order nonlinearities and memory effects by adding further input features such as signal magnitude, phase, or nonlinear power of the signal magnitude, along with their respective delayed versions [10]. By having a richer set of input features, the ARVTDNN was found to require a lesser number of parameters than the RVTDNN and achieve comparable or slightly better results.

From a different perspective, parallel and cascade configurations are basic two-box structures that were frequently used in analytically defined behavioral models to achieve reduced complexity with reliable performance. These architectures divide the system's memory and nonlinear components into discrete processing blocks; they are frequently depicted by models like the Wiener, Hammerstein, or Wiener–Hammerstein forms [1][11]-[13]. This modular representation keeps computational requirements reasonable, simplifies the modeling process, and enables flexible adaptation to different amplifier characteristics [1].

In this work, a behavioral model based on the Hammerstein structure, made of a nonlinear lookup table (LUT) followed by RVTDNN block, is proposed. While the

This work was supported by the Office of Research at the American University of Sharjah under Grant FRG25-E-E63.

RVTDNN is used to mimic the dynamic behavior and memory effects, the LUT is designed to capture the PA's static nonlinear characteristics. This hybrid setup combines the flexibility of NN with the simplicity of analytically inspired structures. It is anticipated that the suggested model will considerably lower the computing complexity in comparison to the ARVTDNN while preserving an equivalent level of modeling accuracy.

The remainder of this article is structured as follows: Section II describes the proposed model and its identification procedure, Section III presents the experimental results as well as a performance comparison with the benchmark model. The conclusions are summarized in Section IV.

II. Proposed Model

The proposed model is devised using a cascaded two-box structure conceptually similar to the Hammerstein model. The main difference is that, in this model, the linear filter commonly used in the Hammerstein model is replaced by a real-valued time-delay neural network (RVTDNN). Compared to advanced neural networks, the RVTDNN has a relatively low complexity since it does not involve additional functions such as those present in long-short term memory (LSTM) based models, and convolutional models.

The block diagrams of the conventional single-box RVTDNN and the proposed model are depicted in Figure 1. For the conventional model, the neural network is fed by the amplifier's input signal (x_{in}) at its input and attempts to predict the measured signal at the output of the amplifier (x_{out}). In the proposed model, the input signal (x_{in}) is initially fed into a memoryless look-up table. The output of the look-up table $(x_{out,LUT})$ is then applied at the input of the RVTDNN model to predict the output signal $(x_{out,est})$ which is ideally a copy of the amplifier's output (x_{out}). Thus, in the conventional single box RVTDNN model, the RVTDNN is used to model the entire behavior of the DUT including the static distortions and the memory effects, and is trained using the measured input and output signals $(x_{in}$ and $x_{out})$. Conversely, in the proposed model, the RVTDNN is solely used to model the residual distortions that are not captured by the LUT. These distortions are mainly made of the memory effects since the nonlinear behavior is often accurately modeled using look-up tables. In this case, the RVTDNN is trained using the look-up table output signal $(x_{out,LUT})$ and the measured output signal (x_{out}).

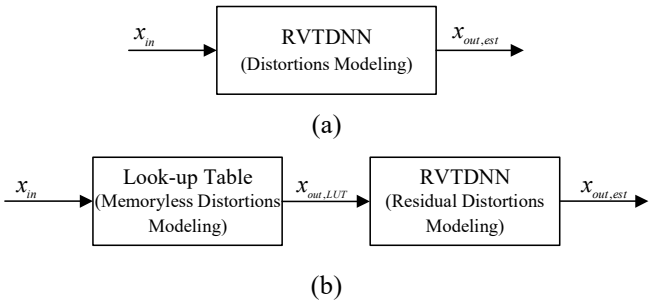


Figure 1. RVTDNN behavioral models. (a) single-box conventional model, (b) proposed two-box model.

The identification process of the proposed model is performed as summarized in the flow chart shown in Figure 2. The measured input and output baseband waveforms of the device under test are acquired and then processed to compensate for the propagation delay and ensure time-alignment between the input and output signals. Then, the measured signals are de-embedded to the input and output planes of the device under test by compensating for any attenuation between the data acquisition plane and the DUT measurement plane. The time-aligned and de-embedded input and output signals are referred to as x_{in} and x_{out}, respectively. These signals are used to synthesize the look-up table. Several approaches can be used for the synthesis of the look-up table including memoryless polynomial fitting and moving average. In this work, the exponentially weighted moving average algorithm was adopted. Once the look-up table is synthesized, the input signal (x_{in}) is applied to it in order to generate the corresponding output signal $(x_{out,LUT})$ which represents the input of the RVTDNN function. This signal $(x_{out,LUT})$ along with the measured output signal (x_{out}) are used for the identification/training of the RVTDNN. The model performance is later assessed by comparing the output of the neural network block $(x_{out,est})$ to the output signal measured at the output of the power amplifier (x_{out}).

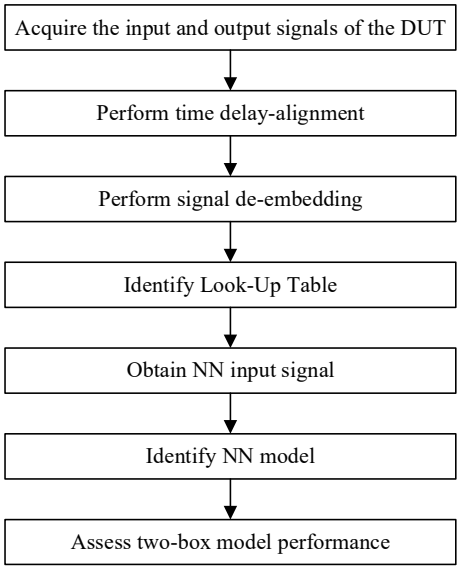


Figure 2. Identification process of the proposed model.

The performance of the proposed model is assessed using the normalized mean-squared error (NMSE) given by

$$NMSE = 10\log_{10}\cdot\left(\frac{\sum_{n=1}^{N}\left|x_{out}(n)-x_{out,est}(n)\right|^2}{\sum_{n=1}^{N}\left|x_{out}(n)\right|^2}\right) \quad (4)$$

where N represents the number of samples in the measured waveforms.

The proposed model is benchmarked against the standalone RVTDNN and ARVTDNN models. The

benchmarking is performed by considering the performance as well as the complexity of each model.

The identification of the look-up table uses linear regression algorithm and therefore its computational complexity can be considered as negligible compared to that associated with the training of the neural network model. Accordingly, it is reasonable to assume that the complexity associated with the model identification is almost that of its neural network part.

The number of neurons per layer and the number of layers were varied in order to evaluate the model performance as a function of its complexity. The model complexity is assessed by computing the total number of parameters used in the neural network (including weights and biases). Figure 3 presents a simplified block diagram of the RVTDNN. The signal at the input of this model is first processed to generate the input features that will used by the model. In the RVTDNN version, the input features consists of the in-phase and quadrature components of the input signal as well as their delayed version. The decision as to how many past terms to include in the input features depends on the memory depth (M) of the device under test being modeled. Therefore, the number of input features (D_0) in the RVTDNN can be expressed as

$$D_{0\_RVTDNN} = 2(M+1) \tag{2}$$

The RVTDNN depicted in Figure 3 contains only one densely connected layer for ease of representation. However, the number of layers can be set to any desired value based on the problem at hand. The number of neurons in each of these layers can be optimized independently. The final layer of the RVTDNN is the output layer which is used to generate the estimated in-phase and quadrature components of the output signal. This output layer contains two neurons. Finally, if needed, a rectangular to polar conversion is performed to generate the complex valued output waveform $(x_{out,est})$.

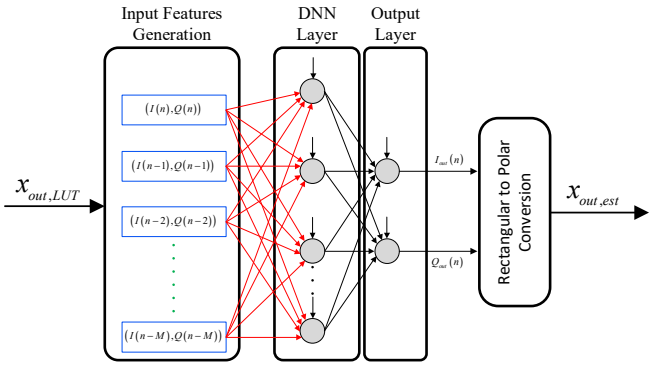


Figure 3. Simplified block diagram of the real-valued neural network.

The total number of parameters in the RVTDNN including weights and biases is given by

$$P = \left[\sum_{l=0}^{L}(D_l+1)\cdot D_{l+1}\right] \tag{3}$$

where L represents the number of layers in the RVTDNN excluding the input and the output layers which correspond to $l=0$ and $l=L+1$, respectively. D_l refers to the number of neurons in layer l.

In the case of the ARVTDNN used for benchmarking, the number of parameters in the model can also be computed using Equation (3). However, the number of input features for a memory depth of M becomes

$$D_{0\_ARVTDNN} = (2+K)(M+1) \tag{4}$$

where K is the largest nonlinearity order of the magnitude of the input signal used as input feature. This is because the input features of the ARVTDNN will be made of the in-phase and quadrature components of the input signal along with their past M values, along with the magnitudes of the input signal $\left(|x_{in}(n)|,|x_{in}(n)|^2,\cdots,|x_{in}(n)|^K\right)$ and their past M values.

III. Testing and Evaluation

A. Device Under Test and Data Acquisition

The device under test used in this work consists of a cascade made of a ZHL-5W-2G-S+ operating as a drive for a custom designed Doherty power amplifier [14]. The power amplifiers lineup was characterized using a 100MHz 5G test signal sampled at 491.52Msps. The carrier frequency was set to 1.425GHz. The test signal has a peak to average power ratio of 10.9dB. The functional block diagram of the experimental setup is presented in Figure 4. The baseband input waveform in downloaded into the RF signal generation board to synthesize the RF signal that will be applied at the input of the driver. The signal at the output of the power amplifier is attenuated and then fed into the receiver that is used to acquire the RF output signal and generate the corresponding baseband in-phase and quadrature components. A Python code was used to process the input and output baseband waveform components to perform the delay estimation and alignment, the data de-embedding to the DUT reference planes, and the model identification and performance assessment.

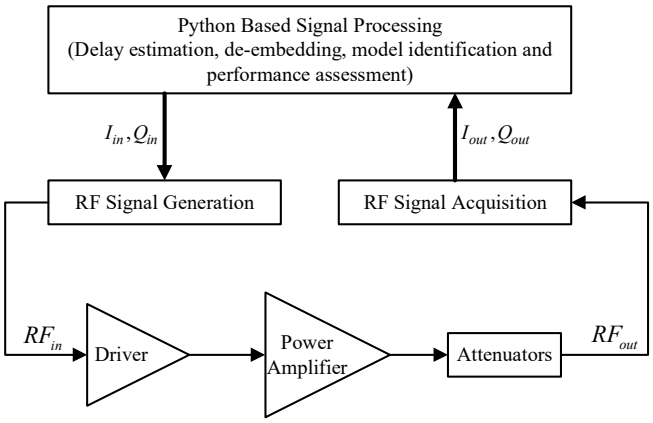
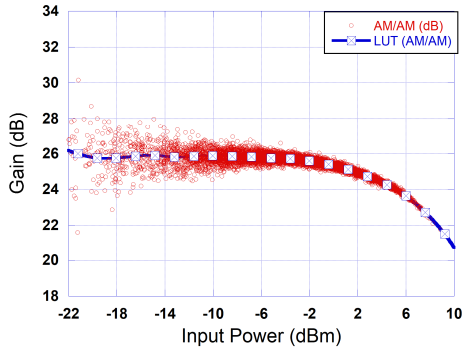


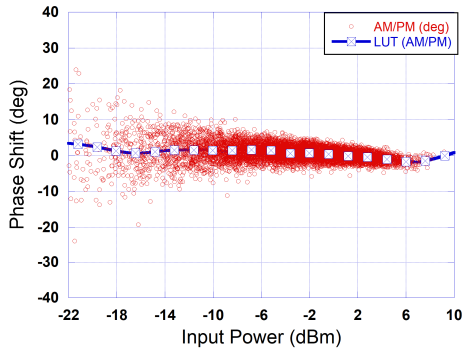
Figure 4. Functional block diagram of the experimental setup.

The AM/AM and AM/PM characteristics of the DUT extracted from the measured input and output waveforms are reported in Figure 5. The AM/AM characteristic shows a pronounced compression region with up to approximately 4dB of compression at peak power. The AM/PM

characteristics show a less pronounced distortion profile. Both the AM/AM and AM/PM characteristics show significant dispersion demonstrating the presence of strong memory effects in the device under test behavior.



(a)



(b)

Figure 5. Measured AM/AM and AM/PM characteristics of the DUT. (a) AM/AM characteristic, (b) AM/PM characteristic.

B. Model Identification and Performance Assessment

First, the LUT function of the proposed model was identified using the exponentially weighted moving average technique. The AM/AM and AM/PM characteristics of the LUT are shown in Figure 5 along with the measured AM/AM and AM/PM characteristics of the DUT. This figure shows that the LUT is able to fit the nonlinear behavior in both characteristics.

For the identification of the neural networks, 70% of the available signals were used for the training, 15% for the testing and 15% for the validation. The activation function used in all layers was the rectified linear unit (ReLU) function. The identification was performed using the adaptive moments (ADAM) optimizer. All neural networks models were trained using 15 epochs.

In this study, it was determined that the memory depth of the DUT is $M = 4$. Thus, the number of input features used in the RVTDNN was 10. The number of layers as well as the number of neurons per layer in the RVTDNN block of the proposed model were varied. For each considered combination the overall performance of the proposed model was assessed in terms of the NMSE. The complexity of the model was evaluated in terms of the total number of parameters in the neural network which was computed for each RVTDNN settings using equation (3). These results are summarized in Table 1 which also includes the performance of the single-box RVTDNN based behavioral model for the same number of layers and neurons. These results show that the proposed model is able to achieve accurate modeling while requiring a very limited number of neurons. In fact, only 106 neurons are needed to achieve an NMSE better than -33dB .

TABLE I. PERFORMANCE AND COMPLEXITY COMPARISION BETWEEN THE PROPOSED MODEL AND THE RVTDNN MODEL

Number of Layers (L)	Number of neurons per layer ($D_1,\cdots,D_L$)	Complexity	NMSE (dB)	
			Proposed Model	RVTDNN
2	(8,2)	106	-33.2	-30.9
2	(16,2)	210	-33.8	-32.8
3	(50,25,2)	1877	-35.4	-35.0
4	(32,18,8,2)	1116	-35.4	-35.1
4	(100,50,25,2)	7477	-35.7	-35.3

The results presented in Table 1 can be explained by the fact that the use of the LUT enables the RVTDNN block to solely focus on the modeling of the residual distortions which are mainly made of mildly nonlinear memory effects. To illustrate this, the AM/AM derived using the signals $x_{out,LUT}$ and x_{out} is reported in Figure 6. For conciseness, the AM/PM characteristic is not shown. However, a similar trend is observed. This AM/AM characteristic represents the behavior that the RVTDNN block of the proposed model attempts to mimic. As it can be seen in this figure, this behavior is mainly linear. Therefore, a limited number of neurons is sufficient to achieve satisfactory modeling accuracy. This also shows that the use of the LUT removes the nonlinear behavior and thus allows the RVTDNN to model the DUT with a low number of parameters without requiring the use of an augmented input features set as it is the case in the ARVTDNN.

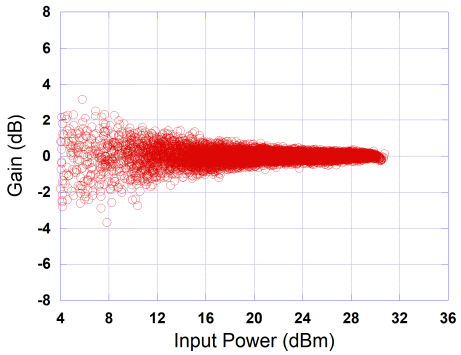


Figure 6. AM/AM characteristic of the residual distortions after applying the LUT function.

To further evaluate the performance of the proposed model compared to state of the art models, the ARVTDNN model was considered. In this model, the input features were generated using the same memory depth $(M = 4)$ as in the RVTDNN models. However, the input features also included the magnitude of the input signal and its powers up to the

TABLE II. PERFORMANCE AND COMPLEXITY OF THE ARVTDNN MODEL

Number of Layers (L)	Number of neurons per layer $(D_1,\cdots,D_L)$	Complexity	Relative Complexity of the Proposed Model	NMSE (dB)
2	(8,2)	226	46.9%	-31.1
2	(16,2)	450	46.7%	-34.4
3	(50,25,2)	1596	69.9%	-33.3
4	(32,18,8,2)	2627	71.4%	-35.6
4	(100,50,25,2)	8977	83.2%	-36.3

third order $(K=3)$. The number of layers and number of neurons per layer were set to the same values as those reported in Table 1 for the RVTDNN networks. The complexity and performances of the standalone ARVTDNN are summarized in Table 2. These results show that at a reduced number of neurons and for two layers networks, the proposed model outperforms the ARVTDNN since it uses a reduced number of features and hence results in a much lower complexity while achieving comparable accuracy. However, for significantly higher complexities (above 2000 parameters), the ARVTDNN has a slightly better accuracy than the proposed model (approximately 0.2dB). Table 2 also shows the relative complexity of the proposed models when compared to the ARVTDNN for the same numbers of layers and neurons per layer. This clearly shows that at low complexity, the proposed model complexity is around 46% that of the ARVTDNN while it achieves comparable NMSE performance.

Figure 7 presents the performance and complexity of the proposed model, the RVTDNN benchmark and the ARVTDNN benchmark for various complexities. These results clearly show that in the region of interest (corresponding to low complexity), the proposed model has the best performance. It is important to mention here that the main objective is to achieve a trade-off between performance and complexity, and that the main factor limiting the adoption of neural networks in field-deployed systems for such applications is their relatively high complexity. Therefore, the proposed model can help resolve this dilemma by allowing neural network based models to achieve better accuracy at a significantly reduced complexity.

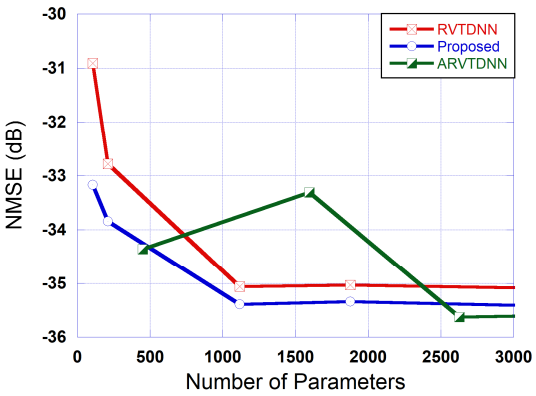


Figure 7. NMSE as a function of the number of parameters for the proposed model and the benchmark models.

A more detailed sweep on the number of layers and the number of neurons was carried out for the proposed model, the RVTDNN, and the ARVTDNN. In this study, the number of input features and the number of neurons in the output layer were kept as defined previously. The number of layers was varied from $L=2$ to $L=4$, and the number of neurons per layer was methodically varied from 5 to 20 in steps of 5. The numbers of neurons per layer were varied such that $D_0 > D_1 \geq D_2 \geq \cdots \geq D_L$. The NMSE results corresponding to this study are reported in Figure 8. This figure only shows the decaying NMSE profile for each model which is obtained by filtering the raw data such that the NMSE is decreasing as the number of parameters increases. This figure corroborates the results previously obtained and clearly demonstrates the superiority of the proposed model in achieving significant performance enhancement at low number of parameters.

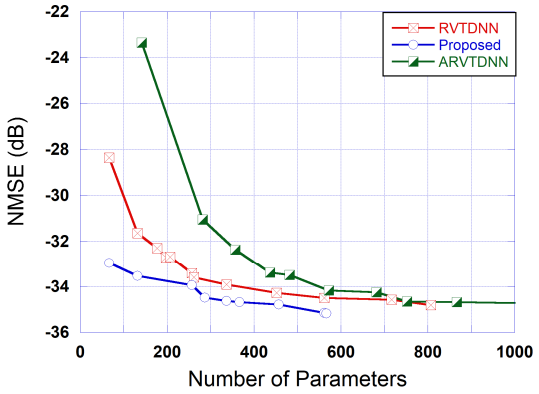


Figure 8. NMSE as a function of the number of parameters for the proposed model and the benchmark models for a sweep on the number of layers and neurons.

IV. CONCLUSION

In this paper, a hybrid two-box model built using the cascade of a look-up table and a real-valued time-delay neural network was proposed of the behavioral modeling of RF power amplifiers. The addition of the LUT upstream of the neural network allows for the modeling of the highly nonlinear behavior of the power amplifier using a simple look-up table. Consequently, the neural network only focuses on the modeling of the residual distortions mainly due to the memory effects. The proposed model was evaluated using experimental data for a Doherty power amplifier driven by a 100MHz 5G test signal. The results show that the proposed model enhances the performance of the standalone RVTDNN especially for a small number of parameters. In fact, for less than 100 parameters in the neural network, the proposed architecture can achieve more than 4dB improvement in the NMSE. Moreover, the use of the LUT/RVTDNN architecture allows for a much lower number of parameters when compared to the ARVTDNN approach. In the performed tests, it was found that the relative complexity of the proposed model, at a reduced number of neurons, is less than 50% of the ARVTDNN while its performance are comparable. The proposed model appears as a viable alternative for hardware friendly RVTDNN structures for power amplifiers behavioral modeling and predistortion.

REFERENCES

[1] F.M. Ghannouchi and O. Hammi, "Behavioral modeling and predistortion," *IEEE Microw. Mag.*, vol. 10, no. 7, pp. 52-64, Dec. 2009. DOI: 10.1109/MMM.2009.934516.

[2] R. Hongyo, Y. Egashira, T.M. Hone, and K. Yamaguchi, "Deep neural network-based digital predistorter for Doherty power amplifiers," *IEEE Microw. Wireless Lett.,* vol. 29, no. 2, pp. 146-148, Feb. 2019. DOI: 10.1109/LMWC.2018.2888955.

[3] S. Zhang, X. Hu, Z. Liu, L. Sun, K. Han, and W. Wang, "Deep neural network behavioral modeling based on transfer learning for broadband wireless power amplifier," *IEEE Microw. Wireless Lett.,* vol. 31, no. 7, pp. 917-920, Jul. 2021. DOI: 10.1109/LMWC.2021.3078459.

[4] T. Liu, S. Boumaiza, and F.M. Ghannouchi, "Dynamic behavioral modeling of 3G power amplifiers using real-valued time-delay neural networks," *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 3, pp. 1025-1033, Mar. 2004. DOI: 10.1109/TMTT.2004.823583.

[5] Q. Xie, Y. Wang, J. Ding, Z. Guo, and J. Niu, "Low complexity convolutional neural network digital predistortion for radio frequency power amplifiers," *in 2024 IEEE 24th International Conference on Communication Technology (ICCT)*, Chengdu, China, Oct. 2024, pp. 1639-1643. DOI: 10.1109/ICCT62411.2024.10946304.

[6] P. Ghazanfarianpoor, S.H. Javid-Hosseini, F. Abbasnezhad, A. Arian, V. Nayyeri, and P. Colontonio, "A neural network-based pre-distorter for linearization of RF power amplifiers," in *2023 22nd Mediterranean Microwave Symposium (MMS)*, Sousse, Tunisia, Oct. 2023, pp. 1-4. DOI: 10.1109/MMS59938.2023.10421055.

[7] S. Li, G. Zhao, C. Yu, F. Li, and Y. Liu, "Power scalable neural network model for wideband digital predistortion," *IEEE Microw. Wireless Lett.,* vol. 33, no. 12, pp. 1658-1661, Dec. 2023. DOI: 10.1109/LMWT.2023.3322273.

[8] S.S. K.C. Bulusu, L. Silva, B. Khan, P. Susarla, N. Tervo, M.J. Sillanpaa, O. Silven, M.E. Leinonen, M. Juntti, and A. Parssinen, "Simplified real-valued time-delay neural network for compensation of power amplifier impairments," *in 2025 IEEE Topical Conf. on RF/Microwave Power Amplifiers for Radio and Wireless Applications (PAWR)*, San Juan, PR, USA, Jan. 2025, pp. 1-4. DOI: 10.1109/PAWR63954.2025.10904050.

[9] A. Ali, and O. Hammi, "Bandwidth, power and carrier configuration resilient neural networks digital predistorter," *IEEE Access*, vol. 11, pp. 63126-63135, Jun. 2023. DOI: 10.1109/ACCESS.2023.3287776.

[10]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. On Neural Networks and Learning Systems.*, vol. 30, no. 1, pp. 242-254, Jan. 2019. DOI: 10.1109/TNNLS.2018.2838039.

[11]O. Hammi and F. M. Ghannouchi, "Twin nonlinear two-box models for power amplifiers and transmitters exhibiting memory effects with application to digital predistortion," *IEEE Microw. Wireless Compon. Lett.*, vol. 19, no. 8, pp. 530-532, Aug. 2009. DOI: 10.1109/LMWC.2009.2024848.

[12]P. Gilabert, G. Montoro, and E. Bertran, "On the Wiener and Hammerstein models for power amplifier predistortion," in Microwave Conf. Proc. (APMC), Suzhou, China, Dec. 2005, pp. 1-4. DOI: 10.1109/APMC.2005.1606491.

[13]T. Liu, S. Boumaiza, and F. M. Ghannouchi, "Augmented Hammerstein predistorter for linearization of broad-band wireless transmitters," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 4, pp. 1340-1349, Apr. 2006. DOI: 10.1109/TMTT.2006.871230.

[14] A.M.E. Abounemra, M. Helaoui, and F.M. Ghannouchi, "Desgin of an efficiency enhanced wideband Doherty Power amplifier based on synthesizing of a modified harmonic-control load modulation network," *IET Microw. Antennas and Propagation,* vol. 18, no. 5, pp. 356-368, May 2024. DOI: 10.1049/mia2.12464.