MACHINE LEARNING-AIDED PIECE-WISE MODELING TECHNIQUE OF POWER AMPLIFIER FOR DIGITAL PREDISTORTION

S. S. Krishna Chaitanya Bulusu[†], *Nuutti Tervo*[†], *Praneeth Susarla*[‡],
Mikko J. Sillanpää[*], *Olli Silvén*[‡], *Markku Juntti*[†], *and Aarno Pärssinen*[†]
† Centre for Wireless Communications (CWC),
‡ Center for Machine Vision and Signal Analysis (CMVS),
∗ Department of Mathematical Sciences (DMS),
University of Oulu, Oulu, FI-90014, Finland.

ABSTRACT

We propose a new power amplifier (PA) behavioral modeling approach, to characterize and compensate for the signal quality degrading effects induced by a PA with a machine learning (ML) aided piece-wise (PW) modeling approach. Instead of using a single pruned Volterra model, we use multiple small-size pruned Volterra models by classifying the input data into different classes. For that purpose, an ML classifier model is trained by extracting some crucial features from both the input signal statistics and the PA operating point. The simulation results indicate that our approach contributes to an improved performance/complexity trade-off than a single generalized memory polynomial (GMP) model in terms of PA behavior modeling and linearization.

Index Terms— Behavioral modeling, computational complexity, decision tree, digital predistortion (DPD), linearization, machine learning (ML), power amplifier (PA).

1. INTRODUCTION

As per green communications obligation, a power amplifier (PA) must be operated as close as possible to its saturation point [1]. However, that leads to wanted effects arising from PA non-linear (NL) behavior such as gain compression, in-band (IB), and out-of-band (OOB) distortions [2].

Digital predistortion (DPD) is a popular PA linearization technique requiring PA behavior modeling [3]. PA is also a NL dynamic system. The Volterra series is computationally expensive but can capture the behavior of PA and memory. It has been theoretically proved that the k-th order pre-inverse of a Volterra system is identical to its post-inverse [4]. This paved way for the usage of pruned Volterra models with reduced complexity to retain most of the modeling capabilities, such as memory polynomial (MP) [5], generalized MP (GMP) [6] and dynamic deviation reduction (DDR) [7] and etc.

ML techniques for DPD have already gained traction [8–11]. However, it is difficult to model the PA behavior with a single model for the entire range of output power because of varying behavior at different power levels. Thus, piece-wise (PW) polynomial-based models have been shown to be quite effective in modeling and linearizing PAs with strong nonlinear effects [12–15]. In [12], a vector switched model has been proposed where the input data samples are classified by a computationally expensive k-means clustering algorithm based on their envelope. Nevertheless, the basic intent of PW modeling and results shown in [12] are interesting as they paved the way with better hardware-friendly techniques such as [13] and [14]. In [13], a low-computation learning algorithm based on a simple decorrelation rule is used for PW modeling. In [14], authors proposed a PW closed-loop DPD solution using low-complexity gradient-adaptive parameter learning algorithms. Recently, a machine learning (ML)-based scheme was proposed in [15] where the input samples are classified by the decision tree classifier using features like current magnitude and past samples. The sub-model coefficients are then extracted.

Our proposed approach involves two stages: ML classification of the input and output baseband signal samples and piecewise digital predistortion (DPD). The novelty of this scheme is to model an accurate and low-complex machine learning (ML)-aided classifier in the first stage involving the input and output baseband signal samples. Here, the boundaries between the classes are calculated as a function of both the input data statistics and PA operating point. In the second stage, the class-wise samples are linearized by tailored GMP models. Supervised ML approaches such as k-nearest neighbors (kNN) and decision trees are non-parametric and do not have an assumption on the distribution of the data. Therefore, the key aspect lies in extracting some essential features from the statistics of the input data samples and also the PA operating point. The simulation results validate this perspective leading to an improved performance/complexity trade-off than a single GMP model in terms of PA behavior modeling and linearization. Albeit, further improvements can be achieved by determining the minimum required number of classes and properly identifying their boundaries.

Rest of the paper is organized as follows. Section 2

This is work was supported in part by the Academy of Finland projects 6Genesis Flagship (grant no 346208) and Profi5 (HiDyn) (grant no 326291).

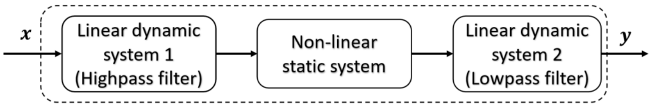

Fig. 1. Wiener-Hammerstein Model of the PA with memory.

presents the simulated PA model and relevant metrics. The proposed scheme is explained in Section 3. The simulation results are given in Section 4 and then the paper is concluded.

2. PA MODEL AND PERFORMANCE METRICS

2.1. PA Model

Let us denote the orthogonal frequency-division multiplexing (OFDM) modulated discrete-time input and output data vectors for PA as $\mathbf{x} = [x_0, \ldots, x_{N-1}]$ and $\mathbf{y} = [y_0, \ldots, y_{N-1}]$, where N is the fast Fourier transform (FFT) size. Then, $\mathbf{x}$ can be mathematically modeled as a function of $\mathbf{x}$ as $\mathbf{y} = \mathbb{F}_a\big(|\mathbf{x}|\big).e^{j(\mathbb{F}_p(|\mathbf{x}|)+\phi_{\mathbf{x}})}$, where, $\mathbb{F}_a$ and $\mathbb{F}_p$ are the classical amplitude-to-amplitude (AM/AM) and amplitude-to-phase (AM/PM) conversion characteristics respectively and $\phi_{\mathbf{x}} = [\phi_0, \ldots, \phi_{N-1}]$ is the phase vector of $\mathbf{x}$. A Wiener-Hammerstein structure [16] is used to model the behavior of a memory PA as illustrated in Fig. 1. A modified Rapp model is used as the NL static system [17]. It is preceded by a high pass filter (HPF) and succeeded by a low pass filter (LPF). The designed LPF is an invertible 3-tap finite impulse response filter with the coefficients $[0.7692, 0.1538, 0.0769]$ [18]. The HPF is designed as an inverse of the LPF.

Let $k \in \mathcal{N}$ be the time index of an input sample x_k where $\mathcal{N} = \{0, 1, \ldots, N-1\}$ is the set of indices of all time samples in $\mathbf{x}$. Then, the mathematical expression of the modified Rapp model's conversion characteristic is

$$y_k = \frac{g_s|x_k|}{\left(1+\left(\frac{g_s|x_k|}{v_{sat}}\right)^{2p}\right)^{\frac{1}{2p}}}.e^{j\left(\frac{\alpha|x_k|^{q_1}}{1+\left(\frac{|x_k|}{\beta}\right)^{q_2}}+\phi_k\right)} \tag{1}$$

where g_s is the small signal gain of PA, p is the knee factor, v_{sat} is the input saturation voltage of PA and α, β, q_1, q_2 are the parameters related to AM/PM characteristic. The input back-off (IBO) of a PA, often expressed in dB, is defined as $\text{IBO} = 10\log_{10}\frac{Nv_{sat}^2}{||\mathbf{x}||_2^2}$, where $||.||_2$ is the Euclidean norm. Likewise, output back-off (OBO) can be defined as the ratio between the mean power of the saturation point of the PA output to the mean power of the output signal.

2.2. Performance and Complexity Metrics

Two metrics are widely used to characterize the quality of the transmitted signal in IB and OOB. These two are error vector magnitude (EVM) and adjacent channel leakage ratio (ACLR) [19]. The ACLR is defined in dB as $\text{ACLR} = 10\log_{10}(\text{P}_{channel}/\text{P}_{adj})$, where $\text{P}_{channel}$ denotes the total power within the assigned channel and P_{adj} is the maximum power of the two adjacent channels. The EVM is measured for the post-amplified signal by using a measurement receiver and is defined both in % and dB as $\text{EVM}_{\%} = ||\mathbf{Y}-\mathbf{X}||_2/||\mathbf{X}||_2 \times 100$ and $\text{EVM}_{dB} = 20\log_{10}(||\mathbf{Y}-\mathbf{X}||_2/||\mathbf{X}||_2)$ and, where $\mathbf{X}$ and $\mathbf{Y}$ are the corresponding complex samples in the frequency domain after the equalization and demodulation of $\mathbf{y}$ and $\mathbf{x}$ respectively.

Computational complexity reduction ratio (CCRR) is used for estimating the proposed scheme's complexity reduction performance over the conventional one. CCRR is defined in % as $\text{CCRR} = \left(1-\frac{C_{new}}{C_{ref}}\right) \times 100$, where C_{new} and C_{ref} are computational complexities of the proposed scheme and the conventional scheme that are used for comparison. We use $F1$ score as the ML classifier performance metric [20].

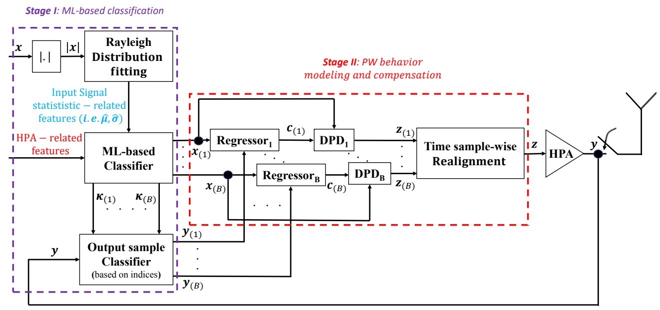

Fig. 2. Block diagram of the proposed scheme.

3. THE ML-AIDED PW-DPD SCHEME

Our proposed scheme, termed as ML-aided PW-DPD, has two stages and follows the direct learning architecture (DLA). It is presented as a block diagram in Fig. 2. Stage I involves ML-based classification of the input data samples into different sample classes using intuitive and easily interpretable supervised learning algorithms like kNN and decision-tree. The selected classifiers are generic and flexible as they are non-parametric and do not assume any inherent distribution of input samples [20]. Under Stage II, we perform behavior modeling using class-wise tailored GMP models followed by DPD for each of the samples. At the end of Stage II, all the class-wise samples are realigned w.r.t. their original time sample indices to construct the predistorted signal $\mathbf{z}$. The mathematical model of GMP is given in [6] and not presented in this paper for the sake of brevity. We denote the GMP model with NL polynomial degree P and memory depth of K for the leading and lagging memories as GMP (P, K) [21].

3.1. Stage I: ML-aided classification

Initially, we scale $\mathbf{x}$ with the IBO coefficient to make Stage I dependent on the PA operating point. In Stage I, the first step is to build essential features for the ML-based classifier of the input data samples. Each input data sample x_k in the baseband may be modeled as a complex Gaussian random variable with Rayleigh envelope distribution if N is sufficiently large [22]. The histogram of $|\mathbf{x}|$ is fitted to a Rayleigh distribution for extracting the estimated statistical parameters such as mean $\hat{\mu}$ and standard deviation $\hat{\sigma}$ based on the maximum likelihood estimation. Now, we need to build some important features for the ML classifier training. Therefore, we extract the four essential features from the statistical properties of the input sample x_k. Those are *square of deviation of the sample's amplitude from mean* (i.e. $(|x_k| - \hat{\mu})^2$), *samples's energy* (i.e. $|x_k|^2$), *real and imaginary parts of the sample* (i.e. $\mathrm{Re}\,|x_k|$ and $\mathrm{Im}\,|x_k|$). We extract the fifth feature from the PA operating point which is the *deviation of the sample's amplitude from saturation voltage* (i.e. $|x_k| - v_{sat}$). These features can be constructed with meager computational complexity.

The ML classifier is trained with these five features and each input sample x_k is classified into one of the B classes. $\forall i \in B$, let $\kappa_{(i)} \in \mathcal{N}$ be the set of indices of the samples that are classified as belonging to the i^{th} class. Then, $\forall \{i, j\} \in B$, we have any two sets $\kappa_{(i)}$ and $\kappa_{(j)}$ being mutually disjoint, i.e. we have $\cup_i \kappa_{(i)} = \mathcal{N}$ and $\cap_i \kappa_{(i)} = \emptyset$. Each output sample y_k will be automatically classified to be the i^{th} class if $k \in \kappa_{(i)}$. In other words, the input and output samples with the same time sample index will belong to the same class but ML classification takes into account only the input sample.

3.2. Stage II: PW behaviour modelling and compensation

We denote classified input and output samples belonging to the i^{th} class as $\mathbf{x}_{(i)}$ and $\mathbf{y}_{(i)}$ respectively. These outputs of Stage I are fed as inputs to the i^{th} Regressor block, Regressor_i in Stage II for the extraction of DPD coefficients for that class which we denote as $c_{(i)}$. All the B Regressor blocks use a tailored GMP model to capture the PA behavior over the samples belonging to the respective classes and may be implemented in a parallel fashion. The proposed scheme also allows for any classical pruned Volterra model or any ML-based model to be used as the regressor in Stage II. In the i^{th} parallel chain, the extracted class-wise DPD coefficients from Regressor_i are fed into the DPD_i block to generate the class-wise predistorted samples $z_{(i)}$.

Finally, all the class-wise predistorted samples are realigned into a discrete-time series as per their sample index positions for the construction of the predistorted signal $\mathbf{z}$ which is then sent to the PA. From the DLA perspective, we identify the PW ideal post-distorter for each class by an improved PA behavior modeling and copying it as the PW predistorter. Our scheme involves only one iteration.

4. SIMULATION RESULTS

We consider 64 quadrature amplitude modulated OFDM symbols $M = 10^5$ and FFT size $N = 1024$ with all active subcarriers and an oversampling factor $L = 4$. The complementary cumulative distribution function of the sample level peak-to-average power ratio of the composite digital waveform, after selected mapping [23], read $8.8\,\mathrm{dB}$ and $9\,\mathrm{dB}$ when measured at 1% and 0.01% points, respectively. The PA gain is normalized in our analysis (i.e. $g_s = 1$) and v_{sat} is $1\,\mathrm{V}$. The modified Rapp model parameters are knee factor $p = 2.25$, $\alpha = -270$, $\beta = 0.17$, $q_1 = 1.1$ and $q_2 = 1.1$. The number of classes in the ML classifier B was chosen to be 5. We have used GMP $(6, 3)$ for the single GMP model and for the proposed scheme's five classes they were chosen by trial and error method to be GMP $(4, 3)$, GMP $(3, 3)$, GMP $(5, 3)$, GMP $(3, 3)$ and GMP $(2, 1)$ respectively.

Table 1. Classwise $F1$-scores over the test dataset for different ML classifier models.

	kNN $F1$ score			DT $F1$ score		
	2 features	4 features	5 features	2 features	4 features	5 features
Class 1	0.05	0.76	**0.95**	0.05	0.95	**1**
Class 2	0.31	0.98	**0.99**	0.31	0.99	**1**
Class 3	0.42	0.99	**0.99**	0.41	1	**1**
Class 4	0.56	0.97	**0.98**	0.53	1	**1**
Class 5	0.02	0.91	**0.95**	0.05	0.98	**0.98**

Our motivation and focus were on the hardware-friendly DPD schemes. It became evident that further improvements can be achieved by determining the minimum required number of the class and properly identifying their boundaries. So, we are of the opinion that it is not fair to compare with the state-of-the-art schemes mentioned in the introduction unless the proposed scheme is fully optimized. Therefore, in this work we do the performance and complexity analysis only with the conventional single GMP model and similar analysis with the aforementioned state-of-the-art schemes shall be included in our future work.

4.1. Performance of ML-based Classification Model

In this subsection, we analyze the performance of ML-based classification using kNN and decision-tree algorithms. The total number of input OFDM signal samples will be MLN. We initially consider a dataset with 8.192×10^6 (i.e. just 2% of the total samples) samples and 5 class labels. In the next step, the aforementioned essential feature information is collected over each sample and used for training/testing the classification algorithms. We note that only 10% of the entire dataset is used for training both kNN and decision-tree algorithms. For the kNN classifier, we chose $k = 10$ and Euclidean distance metric as hyper-parameters.

Table 1 shows the $F1$ score performances of kNN and decision-tree classifiers over the test dataset for all 5 classes, respectively. We notice that classifier models with less than 5 features have shown sub-optimal performance. For both the training algorithms, we observe that if provided with important features then 10% of signal samples is enough to accurately predict the classes for the remaining 90% of the dataset. This shows that, *statistical feature selection favors less algorithmic complexity (training with 0.2% of MLN samples) with reliable classification accuracies*. Moreover,

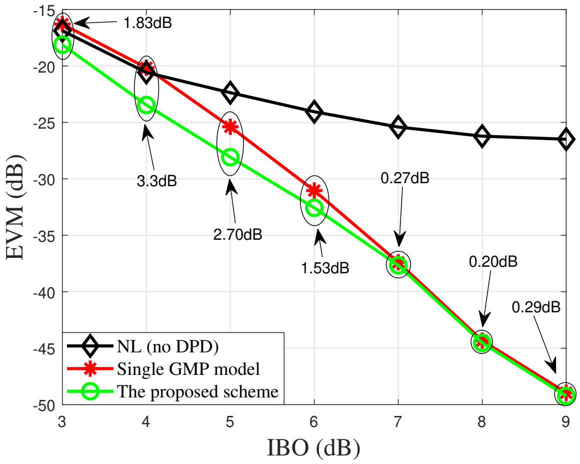

Fig. 3. EVM performance at different IBO values.[1]

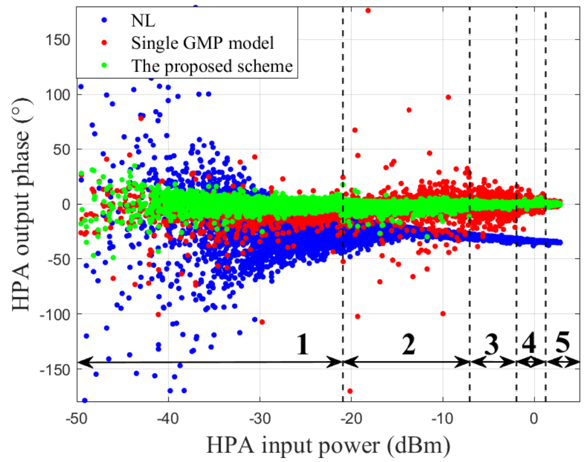

Fig. 5. AM/PM conversion characteristic at 6 dB IBO.[1]

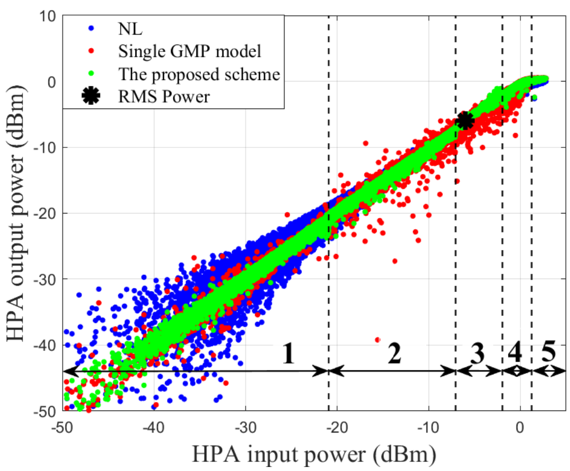

Fig. 4. AM/AM conversion characteristic at 6 dB IBO[1]

Table 2. ACLR Performance in dB at different IBO values.

IBO	OBO	Single GMP model	The proposed scheme	ACLR gain
3	1.52	22.80	24.39	1.59
4	2.18	27.69	30.61	2.92
5	2.91	33.39	34.05	0.66
6	3.69	36.71	36.29	-0.42
7	4.54	38.20	37.89	-0.31
8	5.42	38.75	38.75	0.0
9	6.34	38.79	38.79	0.0

the decision-tree classifier is observed to show the best performance between both ML classifiers with an average $F1$ score over 5 classes ≈ 1.

4.2. Performance and complexity analysis

We analyse the relative EVM improvement (with arrow text) in IB and the relative ACLR gains in OOB for different IBO (and OBO) values as shown in Fig. 3 and Table. 2, respectively.[1] We observe that the proposed scheme outperforms the conventional approach in IB and also for lower IBO values in OOB. The AM/AM and AM/PM plots have been shown at 6 dB IBO in Figures 4 and 5.[1] The extent of each class along with its boundaries is shown where we can see the relative outperformance of the proposed scheme.

At 6 dB IBO, the single GMP model with GMP$(6,3)$ requires 48 coefficients and the proposed scheme requires 113 coefficients. For computational complexity analysis, we try to compare the number of complex multiplications and additions. The number of complex multiplications and additions in a single GMP fitting model is $2MLN\Lambda$ and $2MLN(\Lambda-1)$ respectively, where Λ is the number of DPD coefficients. $\forall i \in B$, the proposed scheme require $\sum_i \big(2\Lambda_i \mathbf{card}(\kappa_i)\big)$ complex multiplications and $\sum_i \big(2(\Lambda_i - 1)\mathbf{card}(\kappa_i)\big)$ complex additions, where Λ_i is the number of DPD coefficients for the ith class and $\sum_i \mathbf{card}(\kappa_i) = MLN$. Compared to the traditional scheme, using CCRR we have estimated 42.44% and 41.56% reduction in complex multiplications and additions for the proposed scheme at 6 dB IBO.

5. CONCLUSION

An ML-aided PW DPD scheme has been proposed by classifying the PA input data based on some crucial features related to the input data's statistics and the selected PA operating point. The OFDM signal dataset built over these features also trained less complex ML classifiers. The simulation results indicate this approach to be promising with an improved performance/complexity trade-off than a single GMP model. However, there is scope for further improvement in the performance. Perhaps the class boundaries may be defined as some function of both the input data statistics and PA behavior instead. Exploring other Volterra models and comparison with state-of-the-art schemes will be part of our future work.

[1] In the legends of all the plotted figures, 'NL' (in black), 'Single GMP model' (in red) and 'The proposed scheme' (in green) indicate conversion characteristic of the original NL PA output (i.e. no DPD), the linearized PA output by the single GMP model and the proposed PW approach, respectively.

6. REFERENCES

[1] I. P. Chochliouros et al., "Energy Efficiency Concerns and Trends in Future 5G Network Infrastructures," in *Energies*, MDPI, 2021, vol. 14, no. 17, pp. 1–14.

[2] L. Guan, and A. Zhu, "Green Communications: Digital Predistortion for Wideband RF Power Amplifiers," in *IEEE Microw. Mag.*, IEEE, 2014, vol. 15, no. 7.

[3] K. C. Bulusu, H. Shaïek, and Daniel Roviras, "HPA Linearization for Next Generation Broadcasting Systems With Fast Convergence-Digital Predistortion," in *IEEE Trans. Broadcast.*, IEEE, 2021, vol. 67, no. 3.

[4] M. Schetzen, "Theory of pth-order inverses of nonlinear systems," in *IEEE Trans. on Circuits Syst.*, IEEE, 1976, vol. 23, no. 5, pp. 285–291.

[5] J. Kim, and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory," in *IEEE Electron. Lett.*, IEEE, 2001, vol. 37, no. 23, pp. 1417–1418.

[6] D. Morgan, Z. Ma, J. Kim, M. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of RF power amplifiers," in *IEEE Trans. Signal Process.*, IEEE, 2006, vol. 54, no. 10, pp. 3852–3860.

[7] A. Zhu, J. C. Pedro, and T. J. Brazil, "Dynamic deviation reduction based volterra behavioral modeling of RF power amplifiers," in *IEEE Trans. Microw. Theory Techn.*, IEEE, 2006, vol. 54, no. 12, pp. 4323–4332.

[8] R. Ma et al., "Machine-Learning Based Digital Doherty Power Amplifier," in *Proc. 2018 IEEE Int. Sym. on Radio-Frequency Integration Tech. (RFIT)*, IEEE, 2018, pp. 1–3.

[9] S. Dikmese et al., "Behavioral Modeling of Power Amplifiers With Modern Machine Learning Techniques," in *Proc. 2019 IEEE MTT-S Int. Microw. Conf. on Hardware and Syst. for 5G and Beyond (IMC-5G)*, IEEE, 2019, pp. 1–3.

[10] T. Wang, W. Li, R. Quaglia, and P. L. Gilabert, "Machine-Learning Assisted Optimisation of Free-Parameters of a Dual-Input Power Amplifier for Wideband Applications," in *Proc. 2019 IEEE MTT-S Int. Microw. Conf. on Hardware and Syst. for 5G and Beyond (IMC-5G)*, Sensors, 2021, vol. 21, no. 8, pp. 1424-8220.

[11] R. Ma et al., "A New Frontier for Power Amplifier enabled by Machine Learning," in *Microwave Journal*, MICROWAVE JOURNAL, 2021, vol. 64, no. 4, pp. 22–32.

[12] S. Afsardoost, T. Eriksson, and C. Fager, "Digital predistortion using a vector-switched model," in *IEEE Trans. Microw. Theory Techn.*, IEEE, 2012, vol. 60, no. 4, pp. 1166–1174.

[13] M. Abdelaziz, L. Anttila, A. Brihuega, M. Allen, and M. Valkama, "Decorrelation-based piecewise digital predistortion: Operating principle and RF measurements," in *Proc. 16th Inter. Symp. Wireless Commun. Syst. (ISWCS)*, IEEE, 2019, pp. 340–344.

[14] A. Brihuega et al, "Piecewise Digital Predistortion for mmWave Active Antenna Arrays: Algorithms and Measurements," in *IEEE Trans. Microw. Theory Techn.*, IEEE, 2020, vol. 68, no. 9, pp. 4000–4017.

[15] Y. Li, X. Wang, J. Pang, and A. Zhu, "Boosted Model Tree-Based Behavioral Modeling for Digital Predistortion of RF Power Amplifiers," in *IEEE Trans. Microwave Theory Techn.*, IEEE, 2021, vol. 69, no. 9, pp. 3976–3988.

[16] P. Crama and Y. Rolain, "Broadband measurement and identification of a Wiener-Hammerstein model for an RF amplifier," in *Proc. 60th Automatic RF Techniques Group (ARFTG) Conf. Dig.*, IEEE, 2002, pp. 49–57.

[17] M. Honkanen and S. G. Haggman, "New aspects on nonlinear power amplifier modeling in radio communication system simulations," in *Proc. 8th Int. Symp. Personal, Indoor, and Mobile Commun. (PIMRC)*, IEEE, 1997, pp. 844–848.

[18] T. Wang and J. Ilow, "Compensation of nonlinear distortions with memory effects in OFDM transmitters," in *Proc. Global Telecommun. Conf. (GLOBECOM)*, IEEE, 2004, pp. 2398–2403.

[19] N. Tervo et al., "Digital predistortion concepts for linearization of mmW phased array transmitters," *Proc. 16th Inter. Symp. Wireless Commun. Syst. (ISWCS)*, IEEE, 2019, pp. 325–329.

[20] A. Géron, "*Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*," 2nd. ed., Sebastopol, CA: O'Reilly, 2019, ch. 3, pp. 94–95.

[21] Mathworks, "Power Amplifier", 2017, `https://se.mathworks.com/help/simrf/ref/poweramplifier.html` [Online].

[22] P. Banelli, and S. Cacopardi, "Theoretical Analysis and Performance of OFDM Signals in Nonlinear AWGN Channels," in *IEEE Trans. on Commun.*, IEEE, 2000, vol. 48, no. 3, pp. 430–441.

[23] K. C. Bulusu, H. Shaiek and D. Roviras, "Potency of trellis-based SLM over symbol-by-symbol approach in reducing PAPR for FBMC-OQAM signals," in *Proc. Int. Conf. on Commun. (ICC)*, IEEE, 2015, pp. 4757–4762.