Behavioral Modeling of a Multimode Power Amplifier Using Perceptrons with AM-AM and AM-PM Complex-Valued Activation Function

Luiza Beana C. Freire[1], Luis Schuartz[1], Edson L. Santos[1], Bernardo Leite[1], André A. Mariano[1], João S. Dias[1] and Eduardo G. de Lima[1]

[1]Electrical Engineering Department of the Federal University of Paraná, Curitiba, Brazil.
e-mail: luiza.chipans@gmail.com

***Abstract*—Three-layer perceptron (TLP) is one of the approaches to the behavioral modeling of radio frequency (RF) power amplifiers (PAs) for wireless communication systems. A low-pass equivalent PA behavioral modeling should be able to represent the amplitude modulation to amplitude modulation (AM-AM) and amplitude modulation to phase modulation (AM-PM) characteristic curves, besides generating only useful contributions. This work shows that adapting a complex-valued activation function to encompass the strict characteristics from the AM-AM and AM-PM curves at the same time fulfilling the bandpass constraint gives better results in terms of network complexity and accuracy. Besides that, a novel TLP-based model for a multimode PA is proposed. This type of PA has characteristics that add difficulty to its inverse modeling. A modification in the ascendant method is also proposed for application with TLP-based PA models. This method simplifies the models in terms of operations needed for implementation. In a scenario of same number of network parameters, a significant reduction in modeling error is achieved when a complex-valued TLP that accounts for both AM-AM and AM-PM is used instead of a complex-valued TLP that accounts only for AM-AM. Such reduction is quantified by up to 5.73 dB (-35.53 dB to -41.26 dB) improvements in normalized mean square error (NMSE) metric. Moreover, when applying the ascendant method a 67% reduction in active parameters is achieved (36 to 12 active parameters), deteriorating the NMSE by less than 0.5 dB.**

***Index Terms*—multimode power amplifier; AM-AM and AM-PM; digital baseband pre-distortion; radio frequency; three-layer perceptron.**

I. INTRODUCTION

Digital baseband pre-distortion (DPD) is currently a linearization strategy widely used for radio frequency (RF) power amplifiers (PAs). The DPD system is designed at baseband level to profit from the benefits of digital signal processing, especially the reduced cost and the capability to generate arbitrary nonlinear mappings with high accuracy [1]. The design of a DPD system requires an accurate and low-complexity nonlinear model for the RF PA [2-4]. The choice of the PA behavioral model is a crucial step in the design of a DPD system. Two classes of dynamic nonlinear models are commonly used: Volterra series and artificial neural networks (ANNs) [5].

In bandpass systems, only contributions at the fundamental zone (in the vicinity of the carrier frequency) can be applied or measured [6]. RF PAs are a special class of bandpass systems. Their behavior can only be modeled by low-pass equivalent terms that comply with some constraints.

This work is an extension of the work published in [7]. In [7] complex-valued three-layer perceptron (TLP)-based RF PA behavioral models are analyzed. The models analyzed were reported in [8], [9] and [10]. They differ because of their complex-valued activation functions. It is shown that less computational complexity in terms of network parameters is presented by the model of [10] in which the real-valued polar angle component passes unchanged through the activation function, in comparison to the models of [8] and [9] that insert it in the sigmoid activation function. This occurs because the complex-valued activation functions used in [8] and [9] generate some useless contributions for the behavioral modeling of RF PAs, whereas the complex-valued activation function used in [10] produces only useful contributions for the behavioral modeling of RF PAs. Avoiding the generation of contributions physically related to harmonic zones can substantially reduce the number of network parameters [11]. Besides, compensating distortions that should not exist is not only a waste of resources but also a source for additional unwanted distortions.

In the activation function reported in [10] only the real-valued amplitude component of the input signal is subject to the nonlinear sigmoid operator and therefore it considers only the amplitude modulation to amplitude modulation (AM-AM) PA behavior. This work proposes a new activation function that takes into account both AM-AM and amplitude modulation to phase modulation (AM-PM) PA behaviors. To our best knowledge, it is the first report that utilizes a complex activation function well fitted to PA behavioral modeling.

Another contribution of this work is to extend the use of neural networks to deal with multimode PAs. In these PAs there are different gain modes, and the choice on which mode the PA should operate depends on the input power level, aiming the highest efficiency. The new activation function approach will be validated using this PA.

Since DPD units need inverse models with the least possible complexity to ensure low energy consumption, the ascendant method for neural networks will be suggested and validated. This method was proposed for models based on memory polynomials and focused on selecting the most relevant parameters to be kept active, reducing the number of operations required in the model [12].

A behavioral modeling simulation is made for the multimode PA described in Sec. II to compare two different models. These two models use a proposed TLP topology

Digital Object Identifier 10.29292/jics.v16i1.223

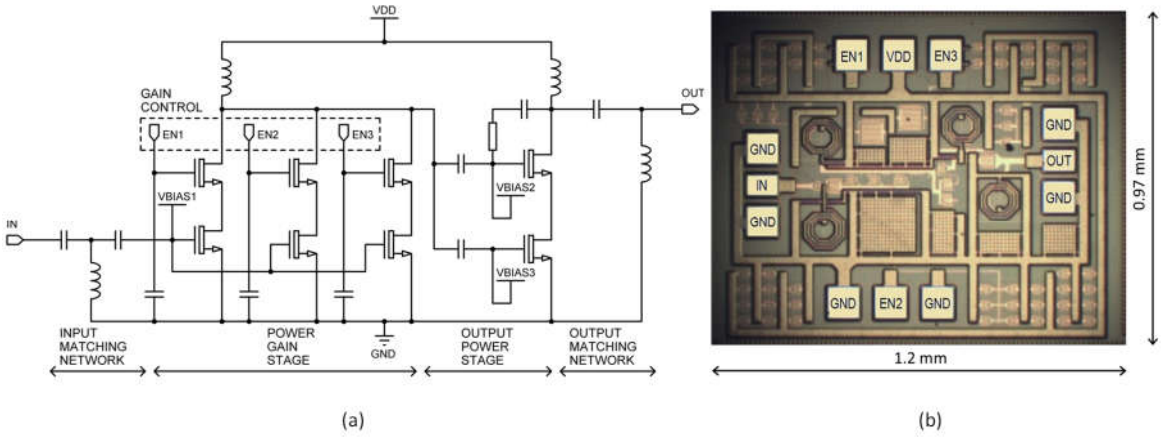

Fig.1 (a) Circuit topology of the PA (bias circuitry is omitted) and (b) chip micrograph of the PA.

designed for the multimode PA and differ on their activation function. One activation function is the proposed one and the other is the one which best performed in [7]. This work is organized as follows: Sec. II details the multimode PA used in this work. Then, the various aspects of the proposed complex TLP-based model for multimode PA is presented in Sec. III. Section IV presents the result analysis from the TLP-based models for multimode PA, besides on the analysis of the proposed ascendant method for simplification of the TLP-based behavioral model. Conclusions are summarized in Sec. V.

II. Multimode PA under Test

The PA is the most critical active component at the transmitter, which consumes a large amount of power in order to support linear transmissions [13]. PAs are typically designed to achieve high performance in the worst operation scenario, i.e., to operate more efficiently at their highest output power. However, high performance is not always required, so that PAs are often over-specified and, consequently, end up consuming more power than necessary. In the case of power-constrained systems, multimode operation may be of particular interest, as it reduces the power consumption. In this paper, we consider a multimode PA proposed in [14]. The PA was designed in Globalfoundries 130 nm CMOS technology to operate in different power gain modes, targeting the 2.4 GHz band. Fig. 1 depicts the designed chip microphotograph and circuit topology, which consists of a reconfigurable power gain stage and a fixed output power stage. The supply voltage is VDD = 1.8 V, whereas that biasing voltages are generated on chip (omitted in Fig. 1). In addition, all impedance matching networks were included on-chip for a low-cost solution. In the gain stage, three parallel cascode cells are adopted for gain control, which can be digitally controlled in six steps through the combination of 3-bits (EN1, EN2 and EN3). In this case, each amplifier cell is enabled with VDD and disabled with GND. Moreover, a stacked topology was adopted at the output power stage to provide sufficiently high transmission power.

Table I. Performance results of the PA for each operating mode at 2.4 GHz.

Mode	EN1	EN2	EN3	Effective Width (μm)	P_{DC} (mW)	OCP_{1dB} (dBm)	Gain (dB)	PAE_{MAX} (%)
1	GND	VDD	GND	80	171.0	13.3	22.4	17.9
2	GND	VDD	VDD	40+80=120	176.4	13.5	25.6	18.4
3	VDD	GND	GND	160	180.0	13.8	27.0	18.6
4	VDD	GND	VDD	160+40=200	187.2	13.7	28.8	18.9
5	VDD	VDD	GND	160+80=240	190.8	13.7	29.8	19.0
6	VDD	VDD	VDD	160+80+40=280	196.2	13.7	31.0	19.0

Measurements of the PA were carried out using on-wafer probing at room temperature. Table I summarizes the measurement results for each operating mode of the proposed PA. An Agilent E83612 vector network analyzer was used to measure the small-signal gain, in which ranges from 22.4 dB in mode 1 to 31 dB in mode 6, so that the power consumption varies from 171 mW (low-gain) to 196.2 mW (high-gain). Notice that when an amplifier cell is enabled, the effective width of the transistors increases, and thus the power gain and the power consumption increase as well. Therefore, when high gain is not required, an amplifier cell can be disabled to save power. The measured large-signal power results were obtained using a Rohde & Schwarz SMF100A signal generator, a Keysight E441B power meter and a Keysight 8485A power sensor. Due to fixed output power stage, the 1 dB output compression point (OCP1dB) is similar for all operating modes. Furthermore, due to high peak-to-average power ratio (PAPR), the PA operates most

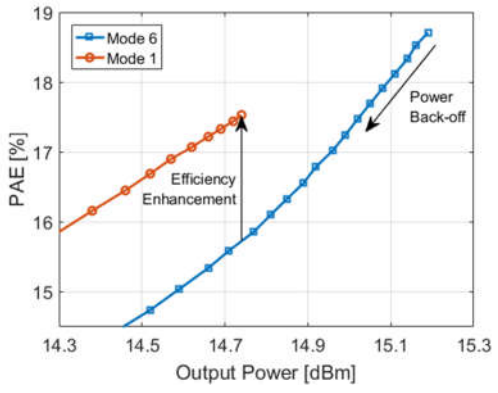

Fig.2 Measured PAE vs. output power.

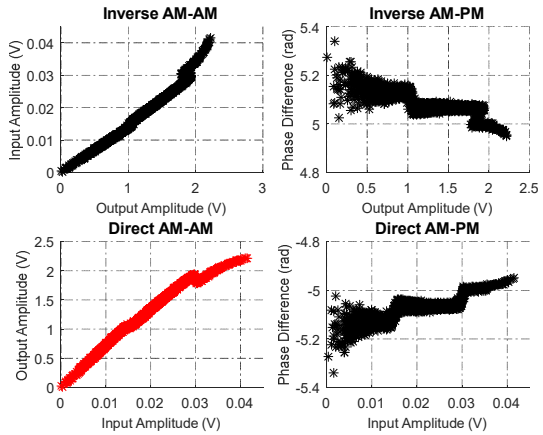

Fig.3 Simulated inverse and direct multimode PA transfer functions.

of the time at power back-off. However, this linear region limits the average output power and, therefore, degrades the efficiency for typical PAs. Another important aspect of the designed multimode PA is to achieve efficiency enhancement at power back-off operation, as illustrated in Fig. 2. From the figure, we can see that in high performance (mode 6), the efficiency decreases as the output power is backed-off. Nevertheless, the PA can switch to mode 1 (low-gain) to improve the efficiency. On the other hand, the PA operating at high power levels, which is more efficient, also has a nonlinear behavior. Therefore, DPD techniques can be used for linearity improvement.

The direct nonlinear PA behavior is mainly described using the output power versus input power curve or AM-AM characteristic and the output minus input phase versus input power curve or AM-PM characteristic. One attribute of the AM-AM curve is its compression characteristic, that describes the change in output power relative to change in input power around the saturation point [15]. An ideal PA should not have interaction between its phase response and power level of the input signal. The AM-PM conversion measures the amount of undesired phase deviation that is caused by amplitude variations of the system.

The simulated AM-AM and AM-PM direct and inverse characteristics of the multimode PA under test are presented in Fig. 3. The PA commutates in real-time among three distinct modes. Each mode actuates in a specific input amplitude interval. The three amplitude intervals have approximately the same length. Although the diverse types of PAs show resemblance characteristics, they are not equal from one another. Examples of other characteristic curves are in [16] and [17]. In this work it will be shown that a TLP that best fits the PA modeling problem is less complex than a generic one.

III. TLP-BASED MODEL FOR MULTIMODE PA

The diagram of the TLP architecture used in this work is shown in Fig. 4. Being E the number of inputs and R the total number of neurons, the E inputs pass through a linear combination with the $w_{r,e}$ weights, where r = 1, 2, ..., R and

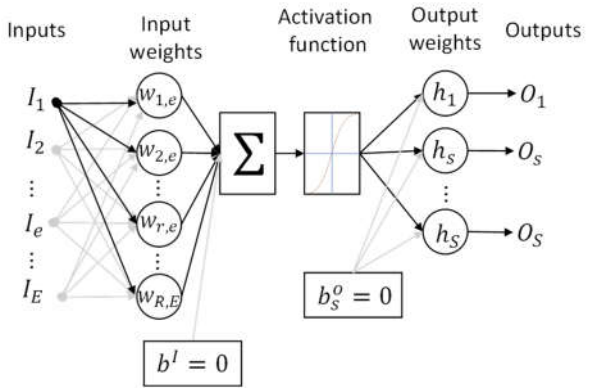

Fig.4 TLP diagram.

e = 1, 2, ..., E. Then the result is subject to the activation function. After that, the signal is multiplied by the output weights h_s, generating the S outputs. For PA modeling, the input bias, b^I, and output bias, b_s^o, are not used and considered equal to zero.

Low-pass equivalent behavioral models deal with complex-valued signals. In this scenario, the TLP architecture can be exploited based on two different strategies. In one case, namely real-valued TLPs, the network inputs, outputs, weights and biases are restricted to be real-valued numbers [6],[18],[19]. Therefore, in real-valued TLPs, the complex-valued signals must be converted into real-valued signals. Alternatively, a complex-valued TLP, having complex-valued inputs, outputs, activation functions, weights and biases, can be used instead [8]-[10]. A complex-valued TLP is going to be used in this work.

This section is going to detail the complex-valued TLP model for multimode PA utilized in this work. First, the influences of the PA signal nature on complex activation functions are discussed. For that a comparative analysis of previously reported complex-valued activation functions is made in Subsection III.A. In Subsection III.B a novel complex-valued activation function that is in resonance with the multimode PA requirements is presented. Subsection III.C discusses how to model a multimode PA based on TLPs. Subsection III.D addresses the simplification of TLP networks by systematic neglect of some parameters based on the ascendant method.

A. Previous complex-valued activation function for TLP-based models.

Different complex-valued TLP-based models are obtained by changing the nonlinear operator that implements the activation function in the hidden layer, Low-pass equivalent models can generate contributions that are not useful for estimating RF PA measurements. Odd-parity with respect to the PA input signal is a necessary but not a sufficient condition for a low-pass equivalent PA model to generate useful contributions. The TLP input biases have constant values that are added to linear combinations of input envelopes. They can easily generate even-order contributions. The generation of even-order contributions can also be attributed to the TLP output bias. Consequently, the odd-parity constraint is satisfied only if all the TLP input

and output biases are null. Therefore, in this work, all the input and output biases of complex-valued TLP-based models are assumed equal to zero.

Keeping the polar angle component multiplied by one is a necessary and sufficient condition for low-pass equivalent RF PA models to generate only useful contributions. Linear operators always preserve the number one that multiplies the polar angle component. They are unable to generate energy at frequencies not stimulated by the generators. Only nonlinear operators can change the number one that multiplies the polar angle component. In complex-valued TLP-based models, just the activation functions F in the hidden layer perform nonlinear operations.

The complex-valued activation functions reported in [8], [9] and [10] are now carefully analyzed. The objective is to investigate if they generate contributions that are not useful for low-pass equivalent RF PA behavioral modeling.

The complex-valued baseband envelope signal related to the PA input is $\tilde{x}(n) = a_n e^{j\theta_n}$, where a_n and θ_n designate the amplitude and polar angle components of $\tilde{x}(n)$, respectively. Then, the input signal of F, $\tilde{u}_r(n)$, in function of the PA input is defined as:

$$\tilde{u}_r(\mathrm{n}) = \sum_{r,e=1}^{R,E} \tilde{x}(n-m) w_{r,e}. \tag{1}$$

In real-valued TLPs, a common choice for F is the odd-parity real-valued sigmoid function:

$$F = \mathrm{tansig}(u_r) = \frac{2}{1+\exp(-2u_r)} - 1, \tag{2}$$

where u_r is a real-valued signal. If (2) is expanded into a Taylor series around $u_r = 0$, then:

$$\mathrm{tansig}(u_r) = \sum_{p=1}^{\infty} \frac{(-1)^{p-1} 2^{2p} (2^{2p}-1) B_p}{(2p)!} u_r^{2p-1}, \tag{3}$$

where B_p is the pth Bernoulli number [20]. Complex-valued TLPs require activation functions capable of handling complex-valued numbers. Three distinct modifications in (2) are reported in the literature. In the first approach, introduced in [8], the exponential operator of (2) can handle complex-valued signals. In this case, the first four terms of the Taylor series expansion of the complex-valued activation function are:

$$\begin{aligned} F = \mathrm{tansig}[\tilde{u}_r(n)] &= \tilde{u}_r(n) - \frac{1}{3}[\tilde{u}_r(n)]^3 \\ &+ \frac{2}{15}[\tilde{u}_r(n)]^5 - \frac{17}{315}[\tilde{u}_r(n)]^7. \end{aligned} \tag{4}$$

All the nonlinear contributions of (4) are useless for estimating RF PA measurements. Any nonlinear contribution of (4) — for any nonlinear order ($3^{rd}, 5^{th}, 7^{th}$) and for any time sample ($n, n-1, \cdots, n-M$) — does not retain one multiplying the polar angle component. For example, if $\tilde{x}^3(n-m)$ is decomposed into real and imaginary parts, then:

$$\tilde{x}^3(n-m) = a_{n-m}^3[\cos(3\theta_{n-m}) + j\sin(3\theta_{n-m})]. \tag{5}$$

In (5), the polar angle component is multiplied by three. In RF PAs, θ_{n-m} is always related to ω_C by $[\omega_C(n-m)+\theta_{n-m}]$. Because $3\theta_{n-m}$ is related to ω_C by $3[\omega_C(n-m)+\theta_{n-m}]$, (5) does not produce any contributions at the fundamental zone.

In the second approach, introduced in [9], the real-valued function of (2) is applied once to the in-band component, $u_r^I(n)$, and once to the quadrature component, $u_r^Q(n)$, of a complex-valued number $\tilde{u}_r(n)$. In this way, the first four terms of the Taylor series expansion of the complex-valued activation function are:

$$\begin{aligned} F &= \mathrm{tansig}[\,u_r^I(n)] + j\,\mathrm{tansig}[\,u_r^Q(n)] \\ &= \left\{u_r^I(n) - \frac{1}{3}[u_r^I(n)]^3 + \frac{2}{15}[u_r^I(n)]^5 - \frac{17}{315}[u_r^I(n)]^7\right\} \\ &+ j\left\{u_r^Q(n) - \frac{1}{3}[u_r^Q(n)]^3 + \frac{2}{15}[u_r^Q(n)]^5 - \frac{17}{315}[u_r^Q(n)]^7\right\}. \end{aligned} \tag{6}$$

Only a subset of the nonlinear contributions of (6) is useful for estimating RF PA measurements. For each nonlinear order ($3^{rd}, 5^{th}, 7^{th}$), (6) generates useful as well as useless contributions. As an example, if $[x^I(n-m)]^3$ is decomposed in amplitude and polar angle components, then:

$$[x^I(n-m)]^3 = a_{n-m}^3\left[\frac{3}{4}\cos(\theta_{n-m}) + \frac{1}{4}\cos(3\theta_{n-m})\right]. \tag{7}$$

Two contributions are reported in (7). In the contribution having the coefficient $3/4$, the polar angle component is multiplied by one. In the contribution having the coefficient $1/4$, the polar angle component is multiplied by three. In order to generate a useful contribution having the coefficient $3/4$, the useless contribution having the coefficient $1/4$ is also generated.

A nonlinear activation function that only generates useful contributions must keep one multiplying the polar angle component. Such requirement is fulfilled by activation functions that are unable to modify the polar angle component. The activation function from [10] processes a complex-valued number without performing any change in the polar angle component. Operators that are unable to change the polar angle component can only modify the amplitude component. The activation function of [10] is described by:

$$F = \{\mathrm{tansig}[|\,\tilde{u}_r(n)\,|]\}\exp[j\angle\tilde{u}_r(n)]. \tag{8}$$

In (8), only the real-valued amplitude component $|\tilde{u}_r(n)|$ is subject to the nonlinear sigmoid operator. The real-valued polar angle component $\angle\tilde{u}_r(n)$ passes unchanged through the activation function given by (8).

The properties of the nonlinear activation function of [10] are now further investigated. The Taylor series expansion of the sigmoid function, reported in (3), can be substituted into (8) in order to get:

$$\begin{aligned} F &= \left[\sum_{p=0}^{\infty} \frac{(-1)^p 2^{2(p+1)}[2^{2(p+1)}-1]B_{p+1}}{[2(p+1)]!} |\tilde{u}_r(n)|^{2(p+1)-1}\right] \exp[j\angle\tilde{u}_r(n)] \\ &= \left[\sum_{p=0}^{\infty} \frac{(-1)^p 2^{2(p+1)}[2^{2(p+1)}-1]B_{p+1}}{[2(p+1)]!} |\tilde{u}_r(n)|^{2p}\right] \tilde{u}_r(n) \\ &= G_{real}[|\tilde{u}_r(n)|^2]\tilde{u}_r(n). \end{aligned} \tag{9}$$

To simplify the notation, in (9) a novel nonlinear operator G_{real} is defined, taking as argument $|\tilde{u}_r(n)|^2$. G_{real} performs a nonlinear mapping between real-valued numbers. Equation (9) provides another interpretation of (8). This activation function can be seen as the product of a complex-valued signal $\tilde{u}_r(n)$ and an even function of its amplitude component $|\tilde{u}_r(n)|$. If only the contributions for $p=0$, $p=1$, $p=2$ and $p=3$ are retained, (9) is reduced to:

$$\begin{aligned} F &= |\tilde{u}_r(n)| \exp[j\angle\tilde{u}_r(n)] - \frac{1}{3}|\tilde{u}_r(n)|^3 \exp[j\angle\tilde{u}_r(n)] \\ &+ \frac{2}{15}|\tilde{u}_r(n)|^5 \exp[j\angle\tilde{u}_r(n)] \\ &- \frac{17}{315}|\tilde{u}_r(n)|^7 \exp[j\angle\tilde{u}_r(n)] \\ &= \tilde{u}_r(n) - \frac{1}{3}|\tilde{u}_r(n)|^2 \tilde{u}_r(n) \\ &+ \frac{2}{15}|\tilde{u}_r(n)|^4 \tilde{u}_r(n) - \frac{17}{315}|\tilde{u}_r(n)|^6 \tilde{u}_r(n). \end{aligned} \tag{10}$$

At this moment, attention is given to the bandpass system whose contributions at ω_C are exactly described by (8). If $\tilde{u}_r(n) = |\tilde{u}_r(n)|\cos[\omega_c n + \angle\tilde{u}_r(n)]$, such bandpass system can be represented by a power series:

$$z_r(n) = \sum_{p=0}^{\infty} c_{2p+1} u_r^{2p+1}(n), \tag{11}$$

in which the coefficients c_{2p+1} are equal to:

$$c_{2p+1} = \frac{\left\{\dfrac{(-1)^p 2^{2(p+1)}[2^{2(p+1)}-1]B_{p+1}}{[2(p+1)]!}\right\}}{\left[\dfrac{1}{4^p}\binom{2p+1}{p}\right]}, \tag{12}$$

for $p = 0,1,2,\cdots,\infty$.

The bandpass system of (11) has the following contributions at ω_C:

$$\begin{aligned} z_r(n) = &\left\{\sum_{p=0}^{\infty} \frac{(-1)^p 2^{2(p+1)}[2^{2(p+1)}-1]B_{p+1}}{[2(p+1)]!} |\tilde{u}_r(n)|^{2(p+1)-1}\right\} \\ &.\cos[\omega_C n + \angle\tilde{u}_r(n)]. \end{aligned} \tag{13}$$

The low-pass equivalent description of (13) is identical to the activation function of (8). Another interpretation of the activation function of (8) is possible. Equation (8) is the low-pass equivalent description that retains just the contributions at ω_C of the bandpass system given by (11). A case study from [7] showed that the complex-valued TLP of [10] presented superior performance when compared with the complex-valued TLPs of [8] and [9].

The limited applicability in low-pass equivalent RF PA behavioral modeling of (4) is due to the fact that it is unable to estimate nonlinearities measured at the RF PA output. Moreover, an increase in modeling accuracy, accompanied by a reduction in the number of network parameters, is expected by using (8) instead of (6), because for each useful nonlinear contribution produced by (6), a useless nonlinear contribution is also generated. For sure, useless contributions will not increase the modeling accuracy. Nevertheless, useless contributions can deteriorate the accuracy of the estimations.

B. New complex-valued activation function for TLP-based PA modeling

A TLP is said real-valued when its weights and activation functions are real-valued, and it processes only real-valued signals. Complex-valued TLPs are capable of processing complex-valued signals and have complex weights and activation functions.

Most models proposed in literature are real-valued and are focused on changing the inputs and outputs presented to the neural networks, as well as their architectures [21]. In recent time, due the trend of deep learning [22], some works utilize new activation functions that emerged [23-26]. Nevertheless, these new activation functions were molded to pattern recognition, natural language processing, recommendation systems, speech recognition, and computer vision applications.

When it comes to complex neural networks there are fewer works, and none focuses on activation functions dedicated to PA modeling [27-28]. Besides that, as sigmoid functions saturate, their use is not indicated with gradient based learning networks [29].

The TLPs, both real and complex-valued, that have an activation function that changes the frequency information of the signal been processed by the network, generate out of band contributions. Then, to ensure that this does not occur, or the inputs must be changed in a way to not present the signal frequency information or the activation function must be adequate to not change it.

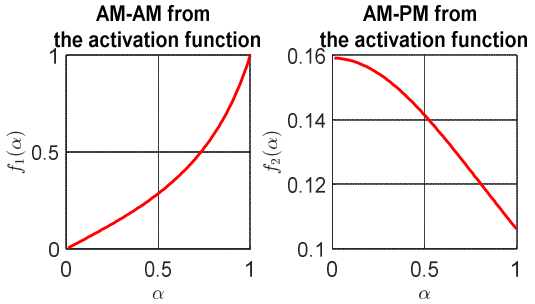

Fig.5 AM-AM and AM-PM characteristics of the proposed activation function of (15). α is the argument of the activation function given by the amplitude of the complex-valued signal applied to the neuron input. f_1 and f_2 are the proposed functions.

When the network inputs are adjusted to meet the requirements mentioned before, it ends up increasing a lot the number of network inputs. So, the benefit of changing the activation function is to diminish the complexity of the network in terms of number of parameters, since it depends on the number of the network inputs. The complexity of the model influences the energy expenditure of the hardware where the DPD unit will be implemented.

Activation functions are mathematical equations that determine the output of a neural network. Besides that, they can insert nonlinear behavior to the network and therefore the capability to model nonlinear systems [18].

In order to best fit the PA behavior and, consequently, diminish the network effort to accurately model it, an activation function should have the capability to encompass all the PA characteristic curves, more specifically the AM-AM and AM-PM behaviors. Also, the activation function should not deform the phase information to not generate out of band contributions. This way the activation function proposed is:

$$F = (f_1(|\tilde{u}|)e^{jf_2(|\tilde{u}|)})e^{j\angle\tilde{u}}, \tag{14}$$

where $\tilde{u}$ is the complex-valued input signal of the activation function F, $|\tilde{u}|$ is its amplitude and $\angle\tilde{u}$ is its phase. Equation (14) is compound of the function f_1 that represents the AM-AM characteristic and of the function f_2 that represents the AM-PM characteristic. Observe that the input angle information is set apart from the nonlinear functions. About the nature of f_1 and f_2 they can take many forms, as long as they are able to represent the PA characteristic curves.

Looking at the inverse AM-AM curve in Fig. 3, certain characteristics can be isolated: the linear behavior with positive slope for lower amplitude values and expansive behavior for greater amplitude values, and its origin in zero. Analyzing the inverse AM-PM curve in same manner, a constant behavior with zero slope for lower amplitudes and a compressive behavior for greater amplitudes can be observed. The activation function does not account for memory effects, instead, past samples of the input signal are inserted in the network to represent this phenomenon.

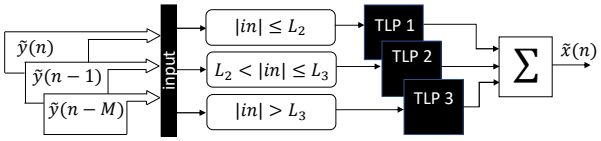

Fig.6 Diagram of the proposed TLP-based inverse model for the multimode PA with three modes.

Inspired on Saleh model [30], the use of rational equations is proposed in here to form the f_1 and f_2 functions. They easily allow to shavpe the curve to present the compressive and expansive behaviors for higher amplitude values, besides the linear lines with zero or positive slopes for lower amplitude values.

One possibility for PA inverse modeling that is proposed in this work is:

$$F = \left[\frac{0.5|\tilde{u}|}{1-0.5|\tilde{u}|^2}e^{j\frac{0.5}{1+0.5|\tilde{u}|^2}}\right]e^{j\angle\tilde{u}}. \tag{15}$$

The proposed activation function of (15) encompasses all the necessary characteristics to properly map the PA inverse transfer function. Fig. 5 shows the shape of the proposed functions f_1 and f_2. The function f_1 initiates at zero, is a line with positive slope for lower input values and has an expansive behavior for larger input values. The function f_2 initiates with a constant value that remains unchanged for lower amplitudes and presents a compressive behavior for larger values. The compresive and expansive behaviors of f_1 and f_2 are dictated by the operator in the denominator. When the compressive behavior is desired the denominator is $1+0.5|\tilde{u}|^2$, otherwise, $1-0.5|\tilde{u}|^2$ is used for the expansive behavior.

C. TLP-based model for multimode PAs

In what concerns the multimode real-time operation, besides of previously presented characteristics, its AM-AM and AM-PM discontinuous curves are a challenge. Modeling the contribution of each mode in a specific TLP is a possible solution [12]. Thus, the block diagram of the TLP-based model for the multimode PA is shown in Fig. 6.

Since the multimode PA presents K operational modes the TLP model presents K networks and K input sets, one for each operational mode. In the end the total signal is recovered by summing the outputs of all TLP networks.

If it comes to direct modeling, then the network inputs are the PA input signals. The K input sets of complex signals are contained within the interval $L_k < |in| \leq L_{k+1}$, where L_k and L_{k+1} are the inferior and superior PA input amplitude limits for each mode, respectively. The amplitude interval values are defined dividing the PA input amplitude range according to its three chosen operational modes, i.e., $k = 1, 2, \ldots, K$. Therefore, $L_1 = 0\,V$ and $L_{K+1} = +\infty\,V$. For this study, since the PA under test is utilized with three modes, K is set equal to 4 and only L_2 and L_3 limits must be determined.

For inverse modeling the network inputs are the PA

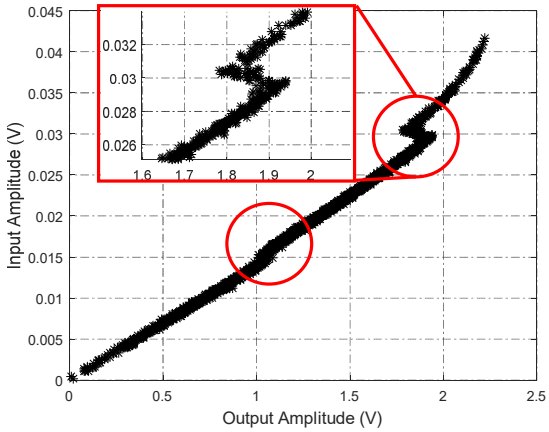

Fig.7 Overlaping of amplitude samples on multimode PA AM-AM inverse behavior.

output signals and some amplitude samples overlap, as shown by the red circles in Fig. 7. That is, for the same output amplitude value there are samples in two modes at the same time, which compromises the 1 to 1 model mapping. So, it is necessary to modify the signal from each mode to exclude these overlapping samples and allow a better delimitation for the transition modes. Considering this issue, a method to define new input intervals for multimode PA inverse modeling that excludes overlapping samples is proposed. In Fig. 8, samples in the first and fourth quadrants overlap. The same thing happens for the second and third quadrants. The samples to be withdrawn are the samples in the even quadrants. The horizontal line is a support to define the inverse L limits.

The excluded samples, N_{exc}, are calculated as:

$$N_{exc} = N - N_m, \tag{16}$$

where N is the total number of samples and N_m is the number

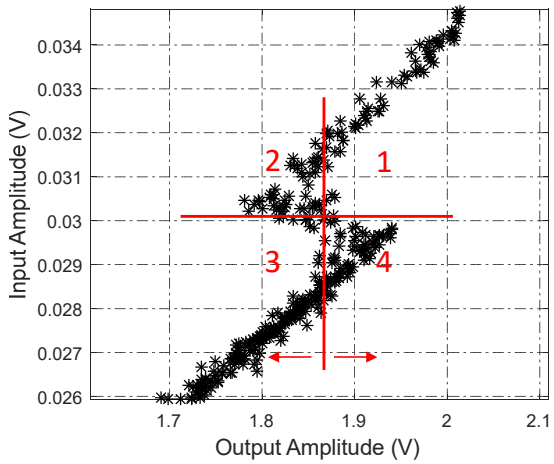

Fig.8 Detailed overlaping samples for the transition between the two modes with higher amplitudes. The vertical line sweeps the values until the L_3 that minimizes the excluded samples is found. The samples to be excluded are in the even quadrants.

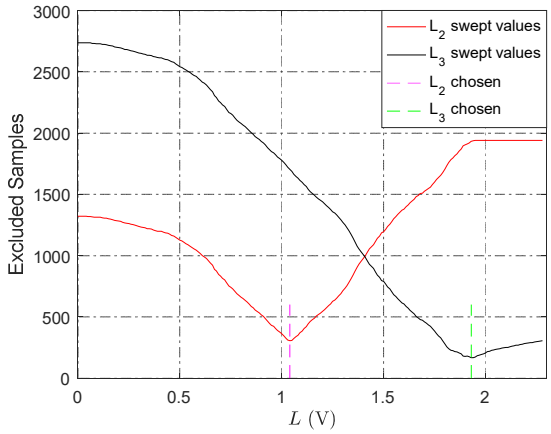

Fig.9 Variation of L_2 in relation to excluded samples with L_3 fixed. Variation of L_3 in relation to excluded samples with L_2 fixed. The valleys indicate the minimum amounts of excluded samples to the respective chosen limits.

of samples from the modified signal.

Those L limits are represented by the vertical lines. The values for the horizontal lines are the L_2 and L_3 limits from the direct modeling (0.015 V and 0.03 V, respectively).

To calculate the new limits for the inverse modeling first L_3 is fixed in a value that includes all samples. So L_3 is set as the maximum amplitude value of the PA output signal. For this PA such value is 2.26 V. Then a sweep is made from 0 to 2.26 V, with 0.01 V intervals, to choose the lower limit L_2 that minimizes N_{exc}.

The same process is made for the upper limit, L_3. First L_2 is fixed in the previous found value. Then a sweep is made from 0 to 2.26 V, with 0.01 V intervals, to choose the upper limit L_3 that minimizes N_{exc}.

Fig. 9 shows the graphs of excluded samples by the swept values. The value chosen for the limit is in the curve valley and indicates the minimum number of excluded samples. After calculating the limits from the modes for the inverse modeling, actual and past samples are applied to the network inputs accordingly as in Fig. 6.

D. Ascendant method

The ascendant method was proposed to memory polynomials models to reduce their amount of parameters [12]. In this work, the ascendant method is employed to simplify the TLP implementation.

Model simplification is realized in the training process. The simplified model keeps the most relevant parameters and cancels the other parameters. Canceling (removing) a parameter in the neural network is the same as forcing it to be equal to zero if it is an input weight ($w_{r,e}$), or to be equal to one if it is an output weight (h_s).

The training process occurs in N_d steps, where N_d is the desired number of active parameters. In each step, n parameters, with $n = 1,2,\dots,N_d$, are activated, and the model performance is assessed each time. The current step utilizes the parameter that gave the model the best performance in the previous step and sets as initial estimative

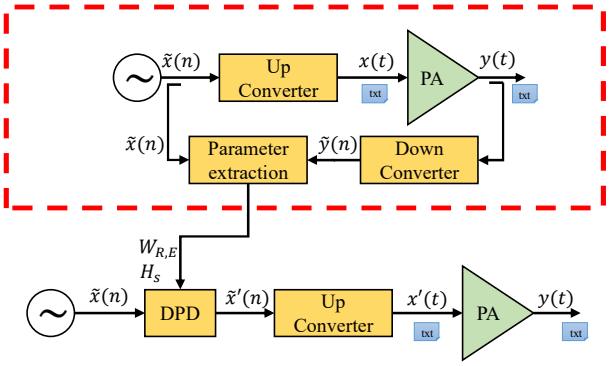

Fig.10 Setup for the PA model extraction (red dashed rectangle) and its use in a DPD unit.

its value from the previous step. At each step a new parameter is elected as the most relevant one, meaning that it is the one that most minimizes the modeling error in that step.

For example, to simplify a model from 10 to 3 active parameters, $N_d = 3$. There will be 3 steps. In the first step all input weights are set to zero and all output weights are set to one. Then one first random parameter is chosen to be activated, the network is trained, and the model performance is assessed. This goes on successively until all parameters are tested one by one. The parameter which best performs will be chosen to go on for the next step. Let us call it w_{r1}. In the second step the network is trained with two parameters. Now, all parameters are canceled beside w_{r1}, and another random parameter is chosen to be activated. w_{r1} starts with its values from the previous step. All the parameters, besides w_{r1}, are tested to choose which would become the relevant parameter from the second step, w_{r2}. For the last step, w_{r1} and w_{r2} are set active, and the last relevant parameter would be chosen as in the previous steps. The initial values for the parameters being tested vary accordingly with the data sets and the method utilized to train the network. In here the value of $1\text{x}10^{-3}$ was used to initiate all parameters that have not been selected in previous steps.

This simplification ends up diminishing the number of operations on complex numbers. When an input weight is set to zero it diminishes one complex multiplication and one complex sum for each input plus one complex sum for each

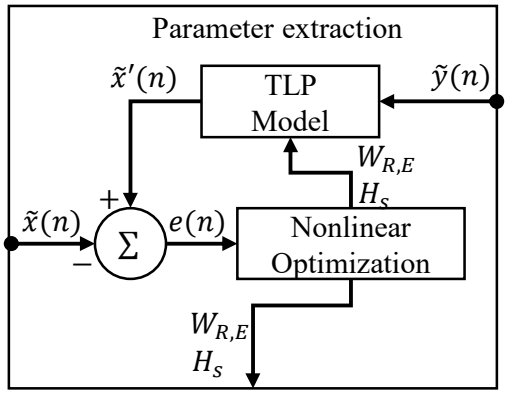

Fig.11 A detailed description of the parameter extraction block defined in Fig. 10.

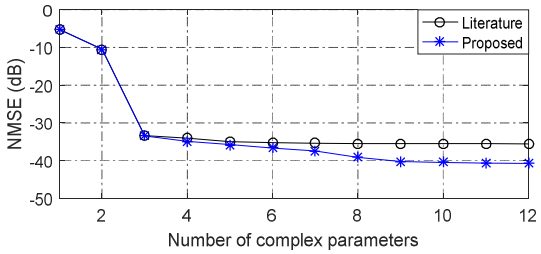

Fig.12 NMSE in function of number of complex-valued parameters for the literature and proposed models with ascendant method. The maximum number of parameters is set to 12.

parameter. When an output weight is set to one it diminishes one complex multiplication. It is important to point out that the number of operations depends on the number of inputs as well as on the quantity of input and output weights that are kept after the simplification.

IV. Simulation Results

In [7] it is shown that the activation function in [10] presents better results if compared with the activation functions from [8] and [9], meaning that it is the best activation function in literature for TLP-based PA modeling. But the activation function in [10] even without deforming the input angle information incorporates only the AM-AM PA behavior and falls short in expectation. In this section, the modeling accuracies of two TLP-based multimode PA modeling approaches are now evaluated to validate the mathematical theory presented in Sec. III. One model uses the sigmoid function as in (8). The other model is the proposed complex-valued TLP, using the activation function of (15).

The setup for obtaining the results is now addressed. Matlab software is used for data processing (modulation, demodulation, parameter extraction and DPD implementation). Transient simulations are done in Cadence Virtuoso software for PA evaluation. Fig. 10 shows the data collection diagram and the PA linearization. The red dashed rectangle indicates the input-output data collection process for later use in the parameter extraction. Yellow blocks are processes made in Matlab and the green blocks represent the PA evaluation in Cadence Virtuoso. The information exchange between software is carried out through text files. Those extracted parameters are used in DPD block.

Table II. NMSE values for proposed and literature models for the corresponding number of complex parameters and without ascendant method.

Number of neurons (R)	Number of complex parameters	NMSE (dB)	
		Proposed	Literature
1	12	-39.29	-35.54
2	24	-41.15	-35.53
3	36	-41.26	-35.53

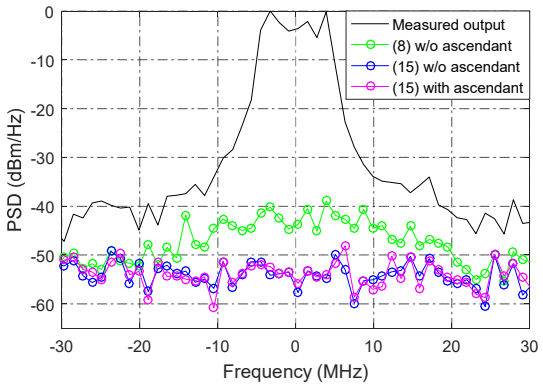

Fig.13 Power spectral densities (PSDs) of the measured output and of the error signals in function of frequency for the TLP-based models been analyzed. The w/o ascendant indicates without ascendant. All models present 36 complex-valued network parameters.

The signal to be amplified ($\tilde{x}(n)$) is complex-valued and has bandwidth and sampling frequency on the order of MHz. This signal is submitted to the up converter block where the resampling and displacement to the carrier occur. The obtained signal ($x(t)$) is real-valued and sampled with a frequency on the order of GHz . This signal is transferred to Cadence Virtuoso through a text file. On Cadence Virtuoso a file in text format with the information to control the switching among the PA modes according to the envelope amplitude is also needed. The Cadence Virtuoso applies a transient simulation with a sampling rate on the order of hundreds of GHz, stimulating and configuring the PA with the data obtained from the text files. The desired signal obtained from the simulation, the PA output voltage for a load of 50 ohms ($y(t)$), is saved in a text file by the Cadence Virtuoso and imported by Matlab. In Matlab, the information passes through the down converter, where the shifting to the baseband, filtering and resampling to MHz occur. Knowing the PA input and output complex envelopes, $\tilde{x}(n)$ and $\tilde{y}(n)$, respectively, the extraction of the model parameters is performed by the parameter extraction block.

The parameter extraction is treated in greater detail in Fig. 11. The TLP model block represents the model described in Fig. 4. For the TLP to mimic the PA inverse behavior, its weights must be adjusted. $W_{R,E}$ and H_S are vectors that contain all the input and output weights, respectively, to be adjusted.

The information obtained from the PA output is applied as input to the model. The model output information, indicated in Fig. 11 by $\tilde{x}'(n)$, is compared to the PA input signal $\tilde{x}(n)$. A nonlinear optimization algorithm is used to adjust the modeling weights. The objective function used in this algorithm is the mean square error.

The TLP-based approaches are applied to the inverse behavioral modeling of the multimode PA described in Sec. II. The RF PA is stimulated by a carrier frequency of 2.4 GHz modulated by an LTE QPSK complex-valued envelope signal having a bandwidth of 10 MHz. The sampling frequency is set to 122.88 MHz. The measured data is divided into two subsets: one for the network training and one for the network validation. Only validation results are reported in the following figures.

The network parameters are identified, in Matlab software, using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton method [31-33]. The memory length is truncated to $M = 2$. The number of neurons in the hidden layer is varied from 1 to 3.

The difference between estimated and measured PA output signals are treated as error signals. The modeling accuracy is analyzed using the normalized mean square error (NMSE) metric, as defined in [34]. For better visualization and understanding of the signals the power spectral densities (PSDs) are also presented [35].

First, the complex-valued TLP-based models without ascendant method are analyzed. Table II shows the NMSE results as well as their corresponding number of network parameters for the different TLP-based PA models. The proposed models, with same number of network parameters, showed substantial reduction in modeling error, quantified by improvements in NMSE metrics up to 5.73 dB.

To exemplify, a NMSE of -41.26 dB is obtained by the TLP-based model using (15) having 36 complex-valued coefficients, whereas the complex-valued TLP model using (8) having the same number of complex-valued coefficients achieved a NMSE of -35.53 dB. Those results confirm the assumptions that it is easier to map the PA behavior when the TLP activation function presents the expansive gain and compressive phase behaviors as depicted in Subsection III.B. Second, the ascendant method was applied to both models, with R = 3 (36 total parameters). The maximum number of active parameters, N_d, was chosen equal to 12. Fig. 12 shows the NMSE results for the various steps of the ascendant method with crescent parameters. The simplified proposed model performed better in terms of NMSE than the simplified literature model. For instance, with 12 active parameters the proposed model achieved a NMSE of -40.80 dB, whereas the literature model obtained a NMSE of -35.53 dB. It should be noticed that the NMSE curve from the literature model stabilized with 5 active parameters, while the proposed model curve kept improving until 12 active parameters. This demonstrates how difficult it was for the literature model to map the PA behavior.

Fig. 13 shows PSDs of the error signals for different TLP-based models. The energy of the error signal is significantly reduced at all frequencies for the proposed complex TLP-based model using (15). In addition, it can be said that the ascendant method is effective, since the energy of the error signals with ascendant method almost overlaps the one without ascendant method, even with a significantly smaller number of active parameters.

V. Conclusions

Two complex-valued TLP-based multimode PA behavioral models were analyzed in this work. The models were applied to the inverse behavioral modeling of a multimode CMOS PA. It was proposed a model that has been proven to generate only in-band contribution and beyond that it uses an activation function that encompasses both AM-AM and AM-PM multimode PA behavior. When compared with another complex-valued TLP-based model which activation function contains only the AM-AM behavior information, it

achieves smaller modeling error for the same number of complex parameters. The ascendant method for neural network models was proposed as well. This method made possible the reduction in 67% (36 to 12 parameters) of the network parameter number maintaining a NMSE of -40.8 dB and hence deteriorating the NMSE by less than 0.5 dB.

We believe this study will be of help to understand the influences of the specific bandpass behavior of RF PAs on TLP activation functions, since it presents a theoretical analysis of several TLP activation functions stimulated by RF PA signals and how to ascertain the generation of only useful contributions. Those knowledges are crucial in the use of TLPs for modeling PAs to guarantee low complexity and high accuracy. Besides that, the ascendant method suggested to simplify the model is one attempt to make it a viable choice in terms of implementation.

ACKNOWLEDGEMENTS

The authors would like to thank the financial support provided by Fundação Araucária de Apoio ao Desenvolvimento Científico e Tecnológico do Estado do Paraná (FAPPR), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

REFERENCES

[1] P. B. Kenington, *High Linearity RF Amplifier Design*, 1st ed. Norwood: Artech House, 2000. ISBN:1580531431

[2] E. B. Solovyeva, "Cascade Structure of Digital Predistorter for Power Amplifier Linearization," *Radioengineering*, vol. 24, 2015, pp. 1071–1076.

[3] H. Rezgui, F. Rouissi, A. Ghazel, "Digital Predistorter Design Using a Reduced Volterra Model to Linearize GaN RF Power Amplifiers," *Radioengineering*, vol. 27, no. 3, 2018, pp. 909–916.

[4] S. Wang, M. Roger, C. Lelandais-Perrault, "Impacts of Crest Factor Reduction and Digital Pre-distortion on Linearity and Power Efficiency of Power Amplifiers*," IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 66, no. 3, 2018, pp. 407–411.

[5] Q. A. Pham, D. López-Bueno, G. Montoro and P. L. Gilabert, "Adaptive Principal Component Analysis for Online Reduced Order Parameter Extraction in PA Behavioral Modeling and DPD Linearization," in *Proceedings of the 2018 IEEE/MTT-S International Microwave Symposium - IMS*, Philadelphia, PA, 2018, pp. 160–163.

[6] X. Peng, X. Qiu, F. Mu, "Digital Harmonic Canceling Algorithm for Power Amplifiers Based on Nonlinear Adaptive Filter," *Prog. In Electromagnetics Research M*, vol. 65, 2018, pp. 151–164.

[7] L. B. C. Freire, L. Schuartz, J. S. Dias, E. G. Lima, "Behavioral Modeling of Power Amplifiers Using Perceptrons with Complex Activation Functions," in *Proceedings of the 35th South Symposium on Microelectronics*, 2020.

[8] N. Benvenuto, M. Marchesi, F. Piazza , et al., "Non linear satellite radio links equalized using blind neural networks," in *Proceedings of the IEEE Int. Conf. Acoust. Speech Signal Processs.*, Toronto, USA, 1991, pp. 14–17.

[9] N. Benvenuto, F. Piazza, A. Uncini, "A neural network approach to data pre-distortion with memory," in *Proceedings of the IEEE Int. Conf. on Communications (ICC)*, Geneva, Switzerland, 1993, pp. 232–236.

[10] G. M. Georgiou, C. Koutsougeras, "Complex domain backpropagation," *IEEE Trans. on Circuis and Systems II: Analog and Digital Signal Processing*, vol. 39, 1992, pp. 330–334.

[11] E. G. Lima, T. R. Cunha, J. C. Pedro, "A Physically Meaningful Neural Network Behavioral Model for Wireless Transmitters Exhibiting PM-AM/PM-PM Distortions," *IEEE Trans. Microw. Theory Tech.*, vol. 59, 2011, pp. 3512–3521.

[12] L. Schuartz, E. L. Santos, B. Leite, A. Mariano, E. G. Lima, "Reduced-Complexity Polynomials with Memory Applied to the Linearization of Power Amplifiers with Real-Time Discrete Gain Control," *Circuits, Systems, and Signal Processing*, 2019, v. 38, pp. 3901–3930.

[13] H. S. Ruiz and R. B. Perez, *Impact of PA on integrated transceivers, in Linear CMOS RF Power Amplifiers: A Complete Design Workflow*, 1st ed. Springer, 2014.

[14] E. L. Santos, M. A. Rios, L. Schuartz, B. Leite, L. Lolis, E. G. Lima, and A. A. Mariano, "A fully integrated CMOS power amplifier with discrete gain control for efficiency enhancement," *Microelectronics Journal*, 2017, vol. 70, pp. 34–42.

[15] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of Communication Systems: Modeling, Methodology, and Techniques*, 2nd ed. Norwell, MA: Kluwer, 2000.

[16] S. Wang, M. A. Hussein, O. Venard, et al., "Impact of the normalization gain of digital predistortion on linearization performance and power added efficiency of the linearized power amplifier," in *Proceedings of the 47th European Microwave Conference (EuMC)*, 2017, pp. 1050–1053.

[17] Q. Hammi, "Augmented twin-nonlinear two-box behavioral models for multicarrier LTE power amplifiers," *The Scientific World Journal*, vol. 14, 2014. Accessed 27 July 2020.

[18] T. Liu, S. Boumaiza, F. M. Ghannouchi, "Dynamic behavioral modeling of 3G power amplifiers using real-valued time-delay neural networks," *IEEE Trans. Microw. Theory Tech.*, 2004, vol. 52, pp. 1025–1033.

[19] M. Isaksson, D. Wisell, D. Ronnow, "Wide-band dynamic modeling of power amplifiers using radial-basis function neural networks," *IEEE Trans. Microw. Theory Tech.*, vol. 53, 2005, pp. 3422–3428.

[20] R. Zucker, *Hyperbolic Functions*. In *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, New York: M. Abramowitz and I. A. Stegun (eds), 1972, pp. 83–86.

[21] F. M. Ghannouchi, O. Hammi and M. Helaoui, *Behavioral modeling and predistortion of wideband wireless transmitters*, John Wiley & Sons, Ltd.: Chichester, 2015, pp. 133–151.

[22] R. Wason, "Deep Learning: Evolution and expansion," *Cognitive Systems Research*, vol. 52, 2018, pp. 701–708.

[23] Z. Liu, X. Hu, T. Liu, X. Li, W. Wang and F. M. Ghannouchi, "Attention-Based Deep Neural Network Behavioral Model for Wideband Wireless Power Amplifiers", *IEEE Microwave and Wireless Components Letters*, vol. 30, no. 1, 2020, pp. 82–85.

[24] R. Hongyo, Y. Egashira, T. M. Hone and K. Yamaguchi, "Deep Neural Network-Based Digital Predistorter for Doherty Power Amplifiers," *IEEE Microwave and Wireless Components Letters*, vol. 29, no. 2, 2019, pp. 146–148.

[25] 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 Transactions on Neural Networks and Learning Systems*, vol. 30, no. 1, 2019, pp. 242–254.

[26] G. C. Tripathi, M. Rawat and K. Rawat, “Swish Activation Based Deep Neural Network Predistorter for RF-PA,” in *Proceedings of the 2019 IEEE Region 10 Conference (TENCON)*, Kochi, India, 2019, pp. 1239–1242.

[27] M. Li, S. He, X. Li, “Complex Radial Basis Function Networks Trained by QR-Decomposition Recursive Least Square Algorithms Applied in Behavioral Modeling of Nonlinear Power Amplifiers,” *International Journal of RF and Microwave Computer-Aided Engineering*, 2019, pp. 634–646.

[28] M. Li, J. Liu, Y. Jiang and W. Feng, “Complex-Chebyshev Functional Link Neural Network Behavioral Model for Broadband Wireless Power Amplifiers”, *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 6, 2012, pp. 1979–1989.

[29] I. Goodfellow, Y. Bengio, A. Courville, Y. Bengio, *Deep learning,* MIT Press: Cambridge , 2016.

[30] A. A. M. Saleh, “Frequency-Independent and Frequency-Dependent Nonlinear Models of TWT Amplifiers,” *IEEE Transactions on Communications*, vol. 29, no. 11, 1981, pp. 1715–1720.

[31] C. G. Broyden, “The Convergence of a Class of Double-Rank Minimization Algorithms,” *Journal Inst. Math. Applic.*, vol. 6, 1970, pp. 76–90.

[32] R. A. Fletcher, “New Approach to Variable Metric Algorithms,” *Computer Journal*, vol. 13, 1970, pp. 317–322.

[33] D. A. Goldfarb, “Family of Variable Metric Updates Derived by Variational Means,” *Mathematics of Computing*, 1970, vol. 24, pp. 23–26.

[34] M. S. Muha, C. J. Clark, A. Moulthrop, et al., “Validation of power amplifier nonlinear block models,” *IEEE MTT-S Int. Microwave Symp. Dig.*, Anaheim, 1999, pp. 759–762.

[35] P. Stoica, R. L. *Moses, Spectral Analysis of Signals*, Pearson Prentice Hall, Upper Saddle River, NJ, 2005