SeMicro-PR 2022

Behavioral Modeling of a PA Using Cascades with Three Volterra Series

Felipe P. Ribeiro, Eduardo G. Lima
Group of Integrated Circuits and Systems (GICS)
Federal University of Parana, Curitiba, Brazil
felipe.pires@ufpr.br, elima@eletrica.ufpr.br

***Abstract*—The linearization of the power amplifier (PA) is an indispensable step in the project of a wireless communication system to improve the PA efficiency and reduce the spread spectrum effect, avoiding interference among neighbor channels. For the linearization of the PA, a behavioral modeling is necessary. This article deals with the study of a cascade among three Volterra series for the behavioral modeling of a class AB PA with GaN-based technology stimulated by a WCDMA signal. The numerical experiments were realized in Python. The parameters were estimated by the separable least squares method (SLS) and optimized through the Levenberg-Marquardt method. All trained models have 250 or fewer coefficients and parameter order truncation of 10. A total of 32284 cases were performed, qualified, treated and presented according to the normalized mean square error (NMSE). The results show that the proposed model, compared to the traditional model, reduces computational complexity and NMSE by up to 51% and -3.787 dB respectively, and compared with the two-blocks Volterra series cascade, reduces computational complexity and NMSE by up to 40% and -1.286 dB respectively. The best NMSE result of the models was -38.83 dB for the traditional model, -40.106 dB for the two-blocks cascade model and -41.359 dB for the three-blocks cascade model.**

***Index Terms*—PA, amplifier, behavioral modeling, Volterra series, cascade model.**

I. Introduction

The Power Amplifier (PA) [1] is the most important device in a wireless communication system, responsible to increase the power of a signal before its transmission, being the major consumer of power and often the most expensive device of the system. Although, its use is compromised by its nonlinear input-output response to high power input levels due to the saturation and compression effects of its internal transistor and the phase distortion caused by capacitors and inductors of the impedance matching networks. Thanks to this, the application of a linearization step is indispensable to improve the PA efficiency and reduce the spread spectrum effect, avoiding interference among neighbor channels.

The Digital Pre-Distortion (DPD) [2] is a widely applied method for the PA linearization purpose, consisting into the previous application of the PA's inverse signal, resulting in a linear relationship between the DPD input and the PA output. For the project of a DPD, a behavioral model of the PA is required, with high accuracy and low computational complexity.

The models employed for the behavioral modeling usually are based on polynomial series with fading memory, like the Volterra series [3], and neural networks, like the time delay neural network (TDNN) [4]. In [5] the cascade among two Volterra series was investigated and better performance was achieved in relation to the traditional model. In this paper, the applied model was an extension of this model: the Volterra series cascade with three concatenated blocks.

The objective of this article was to compare the results of the proposed model with the traditional model and the two-blocks Volterra series cascade model, looking for better results with the expansion of the cascade, keeping constant the maximum number of coefficients. The numerical experiments were performed in Python.

II. Behavioral modeling described by the cascade of three Volterra series

The process of describing the behavior of a device, physically or mathematically, is called *behavioral modeling*. Physical models represent the relations explicitly with linear equations and partial derivative equations, which demand a deep knowledge about the device to be modeled and, consequently, have a significant computational complexity. For the PA modeling purpose, its use is discarded by the principle that we are looking for low-complexity models.

A mathematical modeling consists into the abstraction of the real process, looking at the device like a black box, where only the inputs and outputs are known. The modeling process is related to approximating the so-called "transfer function", which in our model will be represented as the instantaneous gain of the PA, described by:

$$A_v(n) = \frac{|\hat{y}(n)|}{|\hat{x}(n)|} \left[\frac{V}{V}\right] \tag{1}$$

where $\hat{x}(n)$ and $\hat{y}(n)$ are the complex-valued envelopes, respectively, at the input and output of the PA.

In this research, the employed model is based on the Volterra series, a polynomial series with the ability to reproduce memory effects. Its output does not depend only on the instantaneous input, but also on the M previous instants (for the discrete-time model). Thank to this, the output is related to the input in a nonlinear form, being possible to represent the non-linearity and memory effects of the PA, moreover, the series is linear in its parameters without any generality loss.

For the PA purpose, it is convenient to adopt the discrete-time low-pass equivalent Volterra series model [6], described by:

$$\hat{y}(n) = \sum_{p=1}^{P} \sum_{q_1=0}^{M} \sum_{q_2=q_1}^{M} \cdots \sum_{q_p=q_{p-1}}^{M} \sum_{q_{p+1}=0}^{M} \sum_{q_{p+2}=q_{p+1}}^{M} \cdots \sum_{q_{2p-1}=q_{2p-2}}^{M}$$

$$\times \hat{h}_{2p-1}(q_1, \cdots, q_{2p-1}) \prod_{j_1=1}^{p} \hat{x}(n - q_{j_1}) \prod_{j_2=p+1}^{2p-1} \hat{x}^*(n - q_{j_2}) \quad (2)$$

where $(^*)$ denotes the complex conjugate operator, M is the memory length, $P_0 = 2P - 1$ is the polynomial order truncation and $\hat{h}_{2p-1}(q_1, \cdots, q_{2p-1})$ are the low-pass equivalent Volterra kernels. The number of parameters to be determined in (2) is given by

$$L = \sum_{p=0}^{P-1} \frac{[(M+p)!]^2 (M+p+1)}{(M!)^2 (p!)^2 (p+1)} \quad (3)$$

As the number of coefficients in (3) increases quickly with P and M, the method becomes unfeasible for higher orders of M and P, so we are going to investigate another modeling, looking for a complexity reduction. The proposed model, represented by Fig. 1 in a simplified block diagram, is composed by a cascade of three Volterra series in parallel with the PA to be modeled, represented as a black box, where Z^{-1} is the unit delay block, $(M1, P1)$, $(M2, P2)$ and $(M3, P3)$ are the truncation parameters of each Volterra series of the model, $X1$, $X2$ and $X3$ are matrices composed by the product presented in (2) for all index combinations, $H1$, $H2$ and $H3$ are the coefficient vectors of each series, $y_1(n)$, $y_2(n)$ and $y_3(n)$ are the outputs of each block and $E(n)$ is the error between the measured and the estimated output of the PA, $y_3(n)$. The traditional case with only one Volterra series, as the two Volterra series cascade model, are particular cases of the proposed model, respectively when two or one of the blocks has $M = 0$ and $P = 1$ as arguments.

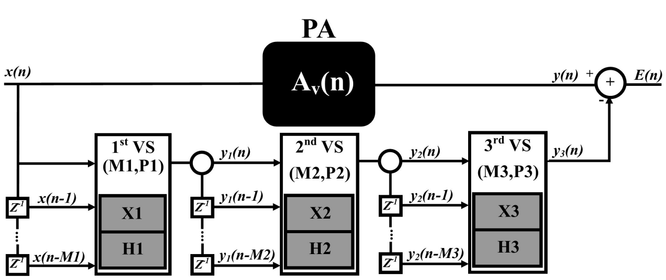

Fig. 1. Proposed model with three Volterra series in cascade

III. Parameter extraction and optimization step

Since the proposed model presents a nonlinear input-output response, its coefficients also present a nonlinear relationship among them, being necessary an optimization step to improve the contributions of all blocks. Commonly the Levenberg-Marquardt method (LM) [7] is employed, which is an algorithm based on gradient decreasing direction search with unconstrained parameters. The method takes at each iteration an approximate of the Jacobian matrix of the system, which tells us the sensitivity of the model in each coefficient, and then takes a direction that decreases the linear approximation of the function (that decrease the error function $E(n)$).

For the LM method, an initial point is required, which can be randomly generated or estimated with a sub-optimal method, like the separable least-squares (SLS) detailed in [5]. Neither methods ensure the convergence to the global optima.

In python, the LM method can be applied through the function *scipy.optimize.least_squares*, which unfortunately does not support complex input or output data functions. For its application, the manipulation of the data is necessary. The objective function optimized in this paper is constructed as:

$$f(HH) = [real\,(E(HH))\,, imag\,(E(HH))]\,,$$

$$HH = [real\,([H1, H2, H3])\,, imag\,([H1, H2, H3])]$$

where HH is the concatenation of the real and imaginary part of all the cascade coefficients, $E(HH)$ is the error when HH is applied to the cascade, and $f(HH)$ is the objective function to be minimized.

The initial guess of the coefficients was taken utilizing the SLS method with the instruction:

$$H\_ = numpy.linalg.pinv(XX\_) @ \hat{y}[M\_ :]$$

IV. Case study for model validation

In this research, the proposed model was applied in a case study with a class AB PA with GaN-based technology, excited by a 900 MHz carrier and stimulated by a WCDMA signal with 3.84 MHz of bandwidth. The signals were measured using a vector signal analyzer (VSA) with a sampling frequency of 61.44 MHz. The data was previously collected and detailed in [8]. For performance analysis, the normalized mean square error (NMSE), was calculated for all computed cases as:

$$NMSE = 10log \frac{\sum_{n=1}^{N} (y_3(n) - y(M1+M2+M3+n))^2}{\sum_{n=1}^{N} y(M1+M2+M3+n)^2} \text{[dB]} \quad (4)$$

where $y(*)$ is the measured output extraction data and $y\_3(*)$ is the estimated output signal, both complex-valued vectors with dimension $N + M1 + M2 + M3$ and N respectively.

For the validation of the method, the model was trained with several combinations of the parameters, that satisfy:

$$0 \leq M1, M2, M3 \leq 10,$$
$$1 \leq P1, P2, P3 \leq 10, \quad (5)$$
$$L1 + L2 + L3 \leq 250$$

where $L1$, $L2$ and $L3$ are the numbers of coefficients respectively of the first, second and third blocks. A total of 32284 cases were performed by the SLS method and then ranked by decreasing NMSE. In Fig. 2 it is presented the scatter plots of the proposed model, evidencing in *A)* the traditional and two blocks cascade model results. We can observe from this figure that the three-blocks cascade Volterra series model was

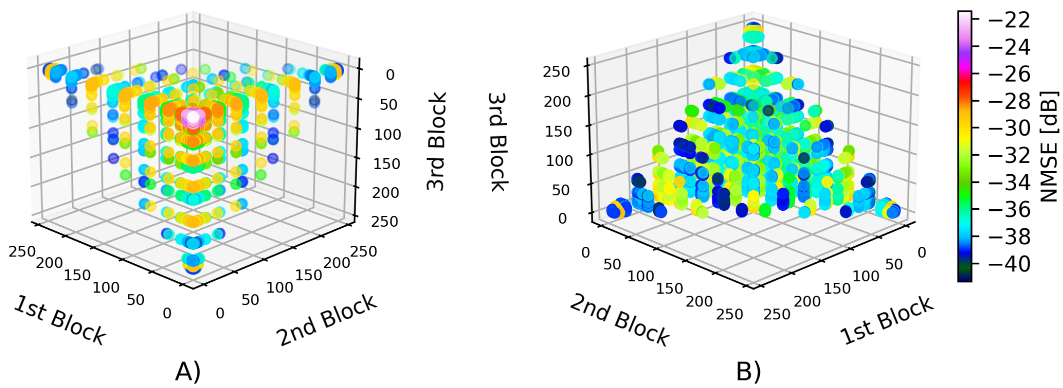

Fig. 2. A) Tetrahedron edges and orthogonal faces represent respectively the results of the traditional model and the two blocks cascade model. B) Tetrahedron interior represent the results of the cascade with three Volterra series

represented in lowest colors temperatures, so, in general, it returned better NMSE results.

Figure 3 presents the NMSE curves for distinct parameter truncation orders, where $T'n'$ represent the curve with truncation order $'n'$. We then observe that the NMSE curves converge for truncation order above to six.

With this assumption, we have filtered the data with the parameter truncation order equal to six, and generated new NMSE curves: three representing the traditional model and the two and three blocks Volterra series cascade models, and three curves of the proposed model considering the number of nonlinear blocks into the scheme (the number of blocks with $M > 0$). All the points in these curves were optimized through the LM method, with the process limited by 100 iterations. The result is presented in Fig. 4. The results show that the proposed model, compared to the traditional model, reduces computational complexity and NMSE by up to 51% (28% on average) and -3.787 dB (-1.218 dB on average) respectively, and compared with the two-blocks Volterra series cascade, reduces computational complexity and NMSE by up to 40% (10% on average) and -1.286 dB (-0.156 dB on average) respectively.

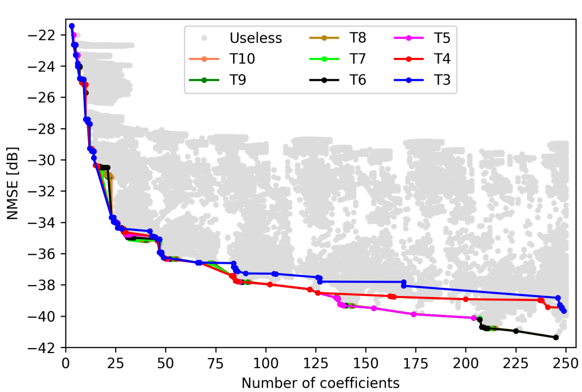

Fig. 3. NMSE curves for different parameter truncation orders

Table I presents the results for the traditional, two-blocks and three-blocks cascade Volterra series models for the extraction and validation datasets applying the best results from Fig. 4B).

Figure 5 presents two plots of these results. The first compares the measured and modeled instantaneous gain of the PA. The second compares the measured and modeled phase shift between the PA output and input. The traditional model

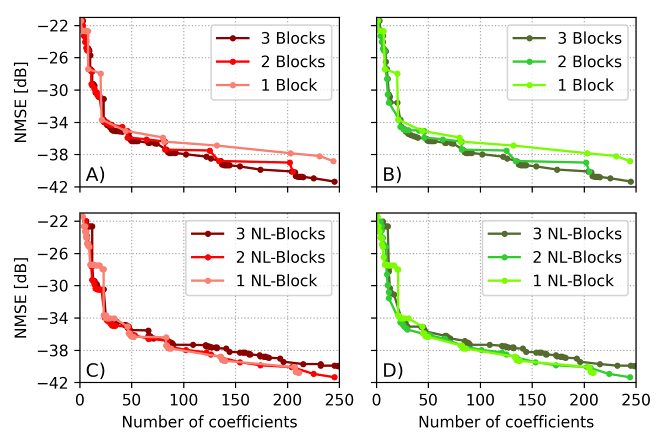

Fig. 4. NMSE curves. A) Traditional model and cascade derivations with two and three blocks (SLS parameter extraction). B) Optimized results from "A)". C) Three blocks cascade Volterra series for different numbers of nonlinear blocks into the scheme (SLS parameter extraction). D) Optimized results from "C)".

TABLE I
COMPARATIVE ANALYSIS OF THE MODELS

M1	P1	M2	P2	M3	P3	L1+L2+L3	NMSE [dB]	
							Extraction	Validation
3	3	0	1	0	1	246	-38.830	-38.233
0	2	6	2	0	1	206	-40.106	-39.544
0	2	6	2	1	4	245	-41.359	-40.595

was excluded from these plots for better visualization. From Fig. 5, there is a smaller dispersion between measured and estimated data when using the model with three blocks in the cascade.

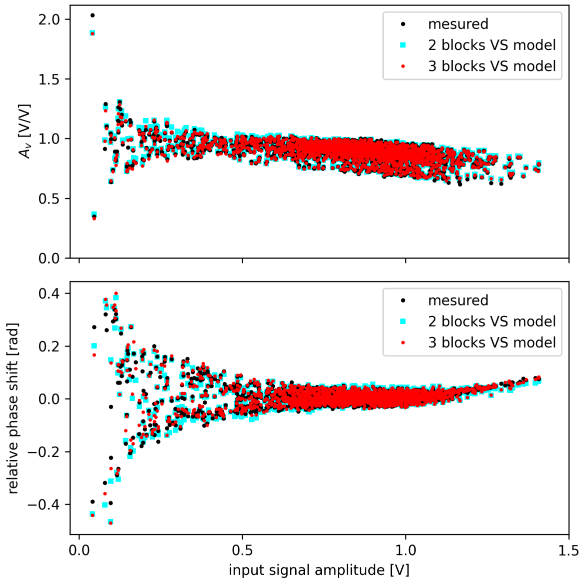

Fig. 5. Measured and modeled instantaneous gain and phase shift

Figure 6 presents the spectrum of the input and output signal applied to the PA and the error spectrum for the three models compared in this paper. From Fig. 6, there is a smaller error in most of the spectrum when using the model with three blocks in the cascade.

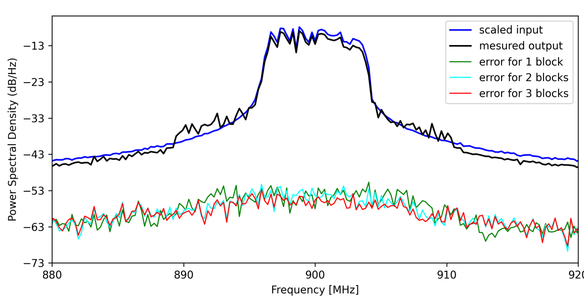

Fig. 6. Power Spectral Density (PSD) for measured input and output signals and modeled errors

Figure 7 presents the error accumulation regions in heatmap graphs for the three compared models. We can observe from this representation that the error converges to zero when the number of blocks in the cascade is increased because the dot scattering decreases. Moreover, as the red region grow up into the third plot we can conclude that more points were close to the "zero error" in relation to the second plot.

V. Conclusions

Through the analysis of the results, we are able to observe the best approaches given by the three-blocks Volterra series

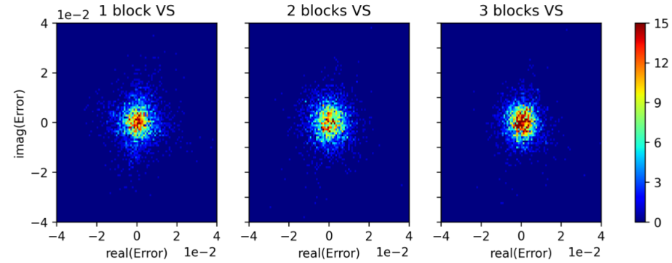

Fig. 7. Heatmap from error accumulation region for the three models

model. The computational complexity was reduced by up to 51% and 40% based respectively on the traditional model and the two-blocks cascade model. The maximum difference between the NMSE curves was -3.787 dB and -1.286 dB in relation, respectively, to the traditional model and the two-blocks cascade model. The error was reduced in the major part of the frequency domain, as presented in Fig. 6, and its converging can be observed in Fig. 7. The set of results denotes the best approaches of the three-blocks cascade Volterra series model. Future works should take into account the optimization based on the module and angle errors and/or global optimization methods.

Acknowledgment

The authors would like to acknowledge the financial support provided by National Council for Scientific and Technological Development (CNPq).

References

[1] H. Wang, K. Sengupta, *"RF and Mm-Wave Power Generation in Silicon"*. Elsevier Science, 2015.

[2] P. B. Kenington, *"High Linearity RF Amplifer Design,"* Norwood. MA: Artech House, 2000.

[3] S. Benedetto, E. Biglieri, and R. Daffara, *"Modeling and performance evaluation of nonlinear satellite links – a Volterra series approach,"* IEEE Trans. Aerosp. Electron. Syst., vol. 15, no. 4, pp. 494–507, Jul. 1979.

[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

[5] F. P. Ribeiro, E. G. Lima, *"Behavioral modeling of radio frequency power amplifiers using cascades of two Volterra series"*, 37th South Brazil Microelectronics Symposium, SIM2022, 2022.

[6] E. G. Lima, *"Behavioral modeling and digital base-band predistortion of rf power amplifiers,"* Ph.D. thesis, Politecnico di Torino, 2009.

[7] J. Dennis Jr., "Nonlinear least-squares," in State of the Art in Numerical Analysis: Conference Proceedings, D. A. H. Jacobs, Ed. London: Academic Press, 1977, pp. 269–312.

[8] A. T. Fukamati, E. G. Lima, *"Modelagem comportamental de amplificadores de potência utilizando cascata de dois polinômios com memória."* In: 2020 Iberchip Workshop, 2020, San Jose, Costa Rica. 2020 Iberchip Workshop, 2020. p. 1-4.