Fixed-Point Radial Basis Function Neural Network for the Digital Baseband Predistortion of an RF Doherty Power Amplifier

A. F. Zanella	R.A.S. Cavalheiro	C. França	L. B. C. Freire	E. G. Lima
GICS (Grupo de Circuitos Integrados e Sistemas)	GICS (Grupo de Circuitos Integrados e Sistemas)	GICS (Grupo de Circuitos Integrados e Sistemas)	GICS (Grupo de Circuitos Integrados e Sistemas)	GICS (Grupo de Circuitos Integrados e Sistemas)
Universidade Federal do Paraná (UFPR)	Universidade Federal do Paraná (UFPR)	Universidade Federal do Paraná (UFPR)	Universidade Federal do Paraná (UFPR)	Universidade Federal do Paraná (UFPR)
andre.felipe.zanella@gmail.com	ricardoacavalheiro@gmail.com	caroldf90@hotmail.com	luiza.chipans@gmail.com	elima@eletrica.ufpr.br

ABSTRACT

This work addresses the digital baseband predistortion (DPD) of radio-frequency power amplifiers (RFPAs) for wireless communication systems. To compensate for the strongly nonlinear and dynamic behaviors observed in RFPAs having a Doherty architecture, in here the DPD is modeled by a radial basis function neural network (RBFNN). Special attention is given to investigate the accuracy of a fixed-point RBFNN DPD as a function of the number of bits used for the representation of the binary numbers. Computer simulations show that, when applied to linearize a Doherty PA behavioral model, an effective fixed-point RBFNN DPD requires at least 25 neurons in the hidden layer and a minimum number of 22 bits for representing the binary numbers. Indeed, if the minimal setup is used, the designed fixed-point DPD improves the ACPR metric at the PA output by 6.6 dB.

Keywords

Digital baseband predistortion, fixed-point arithmetic, power amplifier, radial basis function neural network, radio-frequency.

1. INTRODUCTION

Wireless communication standards for 4G services establish rigorous requirements on linearity [1]. In fact, the high data rates offered by 4G services can only be achieved, in the reduced available bandwidth, if the transmitted RF carrier is modulated by an envelope signal having variable amplitude and with a high ratio between peak and average amplitudes. In this scenario of variable amplitudes, linearity is essential to avoid interference between neighbor users [1]. Wireless communication systems must also provide acceptable power efficiency. Indeed, it is highly desirable to increase the autonomy of the battery present in handsets, as well as to reduce the costs associated with heat dissipation in base-stations [2].

From a microelectronic designer point-of-view, linearity and efficiency are conflicting requirements [2]. This trade-off is accentuated for radio-frequency power amplifier (RFPA) designers. Actually, the RFPAs present at the transmitter chain are based on semiconductor transistors that are subject to an intrinsically trade-off between linearity and efficiency [3]. In other words, traditional RFPAs operating in class A, B or AB can only provide acceptable efficiencies when driven at high power levels, in where non negligible nonlinear distortions are observed.

To improve the power efficiency without deteriorating the linearity, the design of RFPAs for 4G cellular systems are based on the combination of two strategies. In one hand, the RF design uses non-trivial architectures targeting to improve the efficiency. In the other hand, a linearization scheme is added to the transmitter chain.

One example of efficiency enhancement technique for RFPA design is the Doherty architecture [3]. In the Doherty architecture, an auxiliary transistor is associated to the main transistor. While the main transistor is always on, the auxiliary transistor is on only at average to high input power levels. Figure 1 shows an example of a Doherty RFPA transfer characteristic. At low input levels, the output increases linearly with the input. As the input level is increased, a first reduction in gain is observed. At this threshold level, two almost simultaneous phenomena are observed: the gain of the main amplifier starts to compress and the auxiliary transistor turns on. As the input level is further increased, a second gain compression is observed, now due to the compression of the auxiliary transistor. Beyond that point, both main and auxiliary transistors are compressed and the RFPA is close to its saturation level.

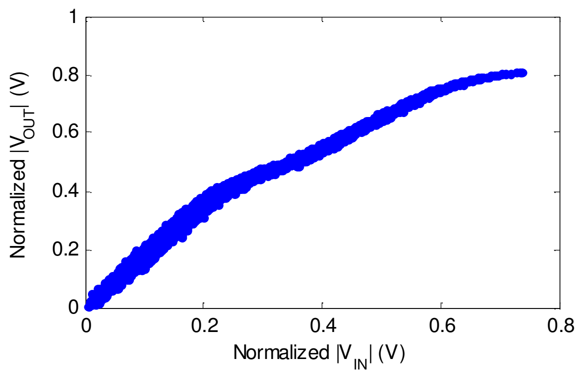

Figure 1. Doherty RFPA transfer characteristic.

Turning the attention to linearization, digital baseband predistortion (DPD) is a cost-effective solution [4]. The DPD is a block connected in cascade with the RFPA, whose purpose is to distort the signal prior to its amplification by the RFPA, so that the signal at the RFPA output is a linear version of the input signal. A fundamental step in the design of a DPD scheme is the choice of a nonlinear dynamic model for the DPD. In literature, Volterra (polynomial) models and neural networks (NNs) are two common alternatives [5]. For the specific case of Doherty PAs, NN is the preferable choice, once the number of parameters in the polynomial models increases very fast with the polynomial order truncation. Indeed, a very high polynomial order truncation is necessary to accurately model strongly nonlinear behaviors like the Doherty transfer characteristic shown in Figure 1. More specifically, in this work the DPD model is given by a NN known in literature as radial basis function neural network (RBFNN) [6].

The contribution of this paper is to investigate an important practical aspect related to the implementation of a fixed-point RBFNN DPD, namely its accuracy as a function of the number of bits used for the representation of the binary numbers. Moreover, the number of neurons in the hidden layer of the RBFNN DPD will be varied in order to estimate its effect on the accuracy of the fixed-point DPD.

This work is organized as follows. Section 2 describes the RBFNN DPD. Section 3 assesses the accuracy of the designed fixed-point RBFNN DPD based on numerical simulations performed on a Doherty RFPA behavioral model. At the end, Section 4 summarizes the conclusions of this work.

2. RBFNN DIGITAL BASEBAND PREDISTORTER

The DPD scheme is based on a cascade connection of the pre-distorter (PD) followed by a power amplifier (PA), as shown in Figure 2. If the PD transfer characteristic has an inverse response in comparison with the PD transfer characteristic, then the signal at the cascade output is a linear version of the signal applied to the cascade input [4], even though the PA is operating at strong nonlinear regimes.

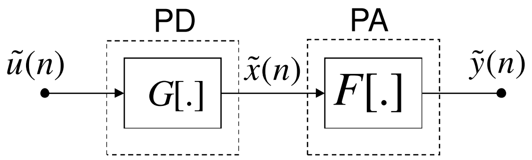

Figure 2. Cascade connection of a PD followed by a PA.

A crucial step in the design of a DPD system is the selection of the DPD topology. In this work, as previously discussed in Section 1, it is chosen the RBFNN architecture [6]. Figure 3 shows the block diagram of an RBFNN having E inputs, R neurons in the hidden layer and S outputs.

In an RBFNN, associated to each one of the R hidden neurons, E parameters called centers (c) are defined. The difference between each applied input and its respective center parameter is performed. Then, the square roots of the sum of the squares of these differences are taken and the obtained results are further multiplied by input bias (b^I) parameters, in order to obtain the signals identified as u in Figure 3. The signals u are applied to activation functions (S), typically Gaussian functions. The O output signals are given by the linear combination of the signals z at the output of the activation functions, having multiplying coefficients designed as (h), and added to output bias (b^O) parameters.

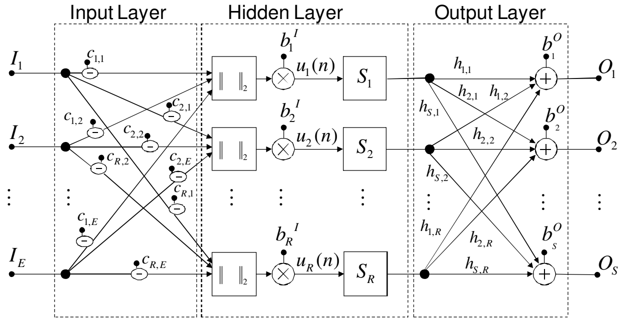

Figure 3. Block diagram of an RBF neural network.

In order to compensate for the memory effects observed at RFPAs, the instantaneous (n) complex-valued envelope at the RFPA output $\tilde{y}$ must be a function of the instantaneous (n), as well as past samples (n-M) up to the memory length M, of the complex-valued envelope at the RFPA input $\tilde{x}$. According to [7], choosing M = 1 and considering that $\tilde{x} = a\exp(j\theta)$ and $\tilde{y} = b\exp(j(\varphi + \theta))$, an RBFNN dealing with just real-valued signals has 4 inputs and 2 outputs, as shown in Figure 4.

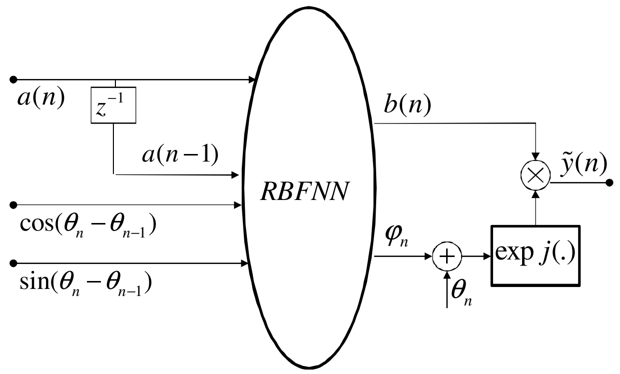

Figure 4. Block diagram of an RBFNN DPD having M = 1.

In order to extract the RBFNN parameters c, b^I, h and b^O, the indirect learning architecture was used [8]. In this algorithm, as shown in Figure 5, it is identified the parameters of a post-distorter (PoD), e.g. an inverse system that is also connected in cascade with the RFPA, put placed after it. Indeed, the parameters of the DPD are just copies of the identified PoD parameters.

Attention is now given to describe an important aspect related to the hardware implementation of the RBFNN DPD shown in Figure 4, namely its accuracy. A first choice that must be done concerns the representation for the binary numbers. While the RBFNN training was done in double-precision floating-point representation, the hardware implementation uses fixed-point arithmetic, in which negative numbers are represented by 2's complement. To convert the floating-point double-precision data (excitation signal and network parameters), to fixed-point data, a routine was created in MATLAB software [9].

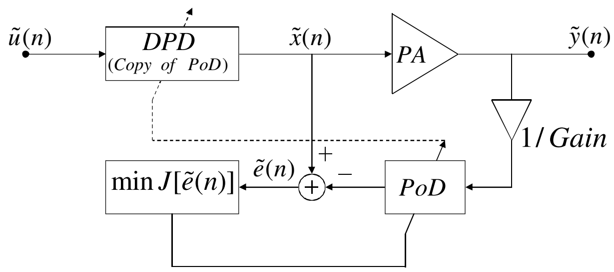

Figure 5. Block diagram of the indirect learning.

According to Figure 3, the RBFNN has R activation functions designated by S, all of them having a nonlinear Gaussian behavior. Besides, a square root operator is previously performed as a step to obtain the signals u applied to the S functions. In the fixed-point arithmetic, nonlinear one-dimensional (1D) operators are implemented by Look-Up-Tables (LUTs). Therefore, to reduce the computational complexity of the fixed-point RBFNN DPD, it suffices to use only R LUTs. In fact, it is enough to use one LUT for each hidden neuron if the LUT is able to take into account the behavior of a cascade between a square root operator and a Gaussian operator.

3. COMPUTER SIMULATIONS

The accuracy of the fixed-point RBFNN DPD described in Section 2 is now assessed based on computer simulations performed on the Matlab software.

The device-under-test (DUT) is a PA behavioral model of a circuit-equivalent GaN HEMT Doherty RFPA, excited by a carrier frequency of 2.14 GHz and modulated by a LTE OFDMA envelope signal having a bandwidth of 10 MHz. The RBFNN architecture of Figure 4 (with R = 20) was also employed for the RFPA behavioral modeling.

The metric adjacent channel power ratio (ACPR) is used here to quantity the accuracy of a DPD. ACPR is given by the power ratios between adjacent and main channels. The ACPR reported here consider a 10 MHz bandwidth for all channels and also a 10 MHz separation between adjacent and main channels. Specifically, the ACPR metric at the PA output is computed for the cases with and without DPD, in a scenario where the RFPA average output powers are the same.

Figure 6 shows the power spectral densities (PSDs) at the RFPA output when the number of neurons in the hidden layer is set to R = 25 and the number of bits used for the representation of the binary numbers is set to 22. Observe that in presence of the DPD, the spectral regrowth is significantly reduced, which means that, in presence of the DPD, the RFPA output signal is more linear in comparison with the case without DPD. This improvement is quantified by an ACPR reduction of 6.6 dB (from -26.9 dB to -33.5 dB), clearly validating the designed DPD as an effective linearizator for the DUT.

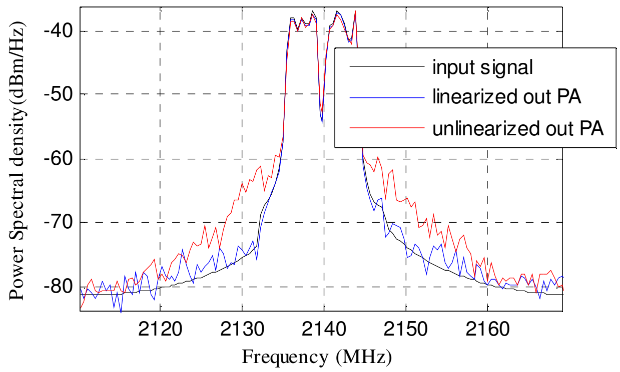

Figure 6. Power Spectral Density, when *R* = 25 and the number of bits used for the representation of the binary numbers is set to 22.

To further validate the designed DPD, in Figure 7 is shown the normalized amplitude of the PA output signal as a function of the normalized amplitude of the OFDMA excitation signal (the AM-AM plot). Observe that a much more linear curve is obtained in the presence of the DPD.

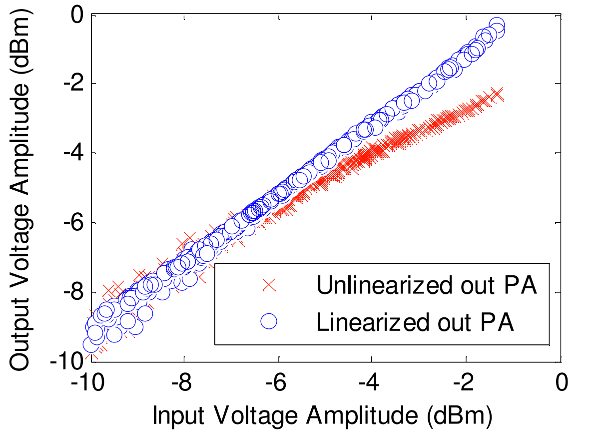

Figure 7. AM-AM plot of unlinearized and linearized PA, when *R* = 25 and the number of bits used for the representation of the binary numbers is set to 22.

At this moment, it will be investigated the impact on the accuracy of the fixed-point RBFNN, if the number of bits used for the representation of the binary numbers is reduced. Figure 8 shows the PSDs at the RFPA output when the number of neurons in the hidden layer is kept equal to R = 25, but the number of bits used for the representation of the binary numbers is reduced to 18. Observe that in this case, the accuracy of the RBFNN is much worse in comparison with the DPD shown in Figure 6 where the number of bits used for the representation of the binary numbers was 22. Therefore, an effective fixed-point RBFNN DPD for the DUT must represent the binary numbers using at least 22 bits

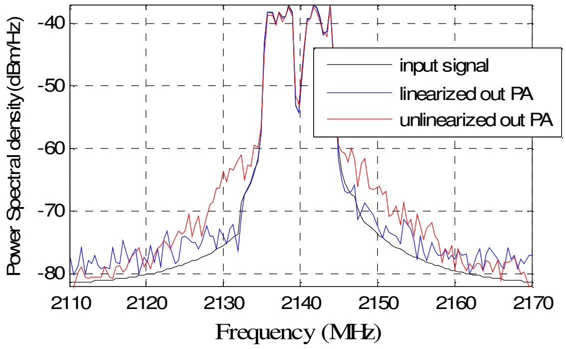

Figure 8. Power Spectral Density, when *R* = 25 and the number of bits used for the representation of the binary numbers is set to 18.

Finally, it will be investigated the impact of varying the number of neurons in the hidden layer on the accuracy of the fixed-point RBFNN. Figure 9 shows the PSDs at the RFPA output when the number of neurons in the hidden layer is set to R = 20 and the number of bits used for the representation of the binary numbers is kept equal to 22. Observe that, in this case, the accuracy of the RBFNN is much worse in comparison with the DPD shown in Figure 6 where the number of neurons in the hidden layer was 25. Therefore, an effective fixed-point RBFNN DPD for the DUT must have at least 25 neurons in the hidden layer.

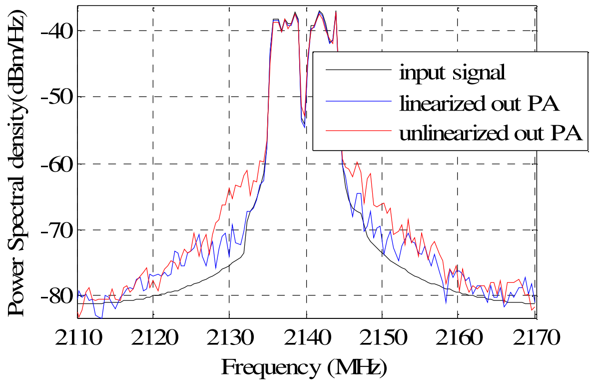

Figure 9. Power Spectral Density, when *R* = 20 and the number of bits used for the representation of the binary numbers is set to 22.

4. CONCLUSIONS

This work has addressed the fixed-point implementation of a digital baseband predistortion having a radial basis function neural network model. The accuracy of fixed-point RBFNN DPDs was investigated based on computer simulations performed on the Matlab software, having a PA behavioral model as the device-under-test. It was reported that an accurate fixed-point RBFNN DPD must have 25 neurons in the hidden layer and the binary numbers must be represented in fixed-point arithmetic having 22 bits. Indeed, 6.6 dB improvements in ACPR metric were achieved by the inclusion of the fixed-point RBFNN DPD, in comparison with the case without DPD. Moreover, drastically deteriorations on the DPD performances were observed, if either the number of neurons in the hidden layer or the number of bits for representing the binary numbers is reduced.

5. ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support provided by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and UPFR/TN under the Program Iniciação em Desenvolvimento Tecnológico e Inovação e Iniciação Científica, UFPR 2013-2014.

6. REFERENCES

[1] D. Raychaudhuri and N.B. Mandayam, "Frontiers of Wireless and Mobile Communications," *Proceedings of the IEEE* , vol.100, no.4, pp.824-840, Apr. 2012.

[2] H. Raab, P. Asbeck, S. Cripps, P. B. Kenington, Z. B. Popovic, N. Pothecary, J. F. Sevic, and N. O. Sokal, "Power amplifiers and transmitters for RF and microwave," *IEEE Trans. Microw. Theory Tech.*, vol.50, no.3, pp.814–826, Mar. 2002.

[3] S. Cripps, *RF Power Amplifiers for Wireless Communications*, 2nd edition. Norwood, MA: Artech House.

[4] P. B. Kenington, *High Linearity RF Amplifier Design*. Norwood, MA: Artech House, 2000.

[5] J. C. Pedro and S. A. Maas, "A comparative overview of microwave and wireless power-amplifier behavioral modeling approaches," *IEEE Trans. Microw. Theory Tech.*, vol. 53, no. 4, pp. 1150–1163, Apr. 2005.

[6] S. Haykin, *Neural Networks: a Comprehensive Foundation*, 2nd edition. New Jersey: Prentice Hall, 1999.

[7] E. G. Lima, T. R. Cunha, and 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, no. 12, pp. 3512–3521, Dec. 2011.

[8] C. Eun and E. J. Powers, "A new Volterra predistorter based on the indirect learning architecture", *IEEE Trans. Signal Process.*, vol. 45, no. 1, pp. 223–227, Jan. 1997.

[9] *MATLAB*. The MathWorks Inc., Natick, MA. http://www.mathworks.com/help/matlab/.