Incremental DPD Linearization for Mobile Terminals with Non-Flat Frequency Response in Dynamic Bandwidth Re-Allocation Scenarios

Wantao Li[#1], Yan Guo[$2], Gabriel Montoro[#], Pere L. Gilabert[#]
[#]Department of Signal Theory and Communications, Universitat Politècnica de Catalunya, Spain
[$]Wireless Terminal Chipset Algorithm Department, Huawei Hisilicon, China
[1]wantao.li@upc.edu, [2]guoyan16@huawei.com

***Abstract* — This paper presents a low-complexity open-loop digital linearization system for handset applications with non-flat frequency response when considering a 5G new radio dynamic bandwidth re-allocation scenario. The proposed digital predistorter (DPD) is scalable with the signal bandwidth by simply activating or deactivating basis kernels, without re-training nor re-assigning the model coefficients. To support this feature, an incremental bandwidth (IBW) generalized memory polynomial behavioral model is proposed, and the coefficients adaptation procedure is described. In order to find the most relevant basis supporting the IBW feature, a constrained version of the doubly orthogonal matching pursuit (DOMP) algorithm is proposed. Experimental results considering a handset power amplifier (PA) SoC under a mismatched load condition (i.e., non-flat frequency response) will validate the proposed approach.**

***Keywords* — power efficiency, linearization, digital predistortion, dynamic bandwidth, power amplifier.**

I. Introduction

In 5G new radio (NR), spectrum resources can be flexibly assigned at different frequency locations and with different bandwidths. It is therefore possible real-time reassignment of both the location and the number of resource blocks (RBs are groups of 12 contiguous subcarriers in the frequency domain), according to the application or radio environment. This requires some degree of re-configurability or the design of robust open-loop solutions capable of coping with the changing transmission requirements.

Power efficient amplification is required in several frequency bands taking into account multiple numerology and channel bandwidths. For narrow-band signals, memoryless digital predistortion (DPD) linearization may be enough to address the PA unwanted nonlinear behavior. Instead, for wideband signals, a more computationally complex DPD model is required to cope with the typical non-flat frequency response (due to load, antenna mismatch) in handsets [1]. Therefore, new challenges arise in the design of the DPD when the PA gain and frequency response change with the bandwidth and frequency of operation. In this context, the desired DPD has to contemplate: i) scalability, the DPD model should allow dynamic sizing to cope with different bandwidth configurations, preferably, by only activating or deactivating basis kernels; ii) adjustability, the DPD coefficients should re-tunable when the frequency location changes to cope with the different PA gain responses; iii) robustness, the DPD should work well in open-loop (i.e., without adapting the DPD coefficients) for all possible RB configurations and locations considering a non-flat frequency response at the PA output due to the unwanted load (antenna) mismatch.

In this paper, by considering a strong load-mismatch condition of the mobile terminal PA, we propose an incremental bandwidth (IBW) DPD model capable of linearizing the PA when operated with different signal bandwidth configurations, by simply activating or deactivating basis kernels. The IBW-DPD assumes that it is possible to append independent DPD basis to cope with the increasing (with the signal bandwidth) PA distortion. A constrained version of the doubly orthogonal matching pursuit (DOMP) algorithm [2] is proposed to incrementally select the most relevant basis. The test signals are 5G NR signals in a 100 MHz channel with 30 KHz sub-carrier spacing and a maximum number of RBs, $N_{\mathrm{RB,max}} = 273$. The signal bandwidth is determined by the number of RBs, $N_{\mathrm{RB}} \in [1, 273]$, and the centering frequency is determined by the center RB location ($L_{\mathrm{RB}} \in [0, 272]$). Fig. 1 shows all possible RB configurations (in terms of N_{RB} and L_{RB}) falling in the shaded triangle area. The lower the N_{RB}, the more possible RB locations we can have. On the other dimension, the higher the N_{RB}, the fewer L_{RB} configurations are available. In this work, we fixed the center RB location at $L_{\mathrm{RB}} = 136$, and focused on the scenario with dynamic N_{RB} (i.e., bandwidth) re-allocation.

II. IBW GMP Linearization Model

The dashed lines in Fig. 1 show the frequency response for different N_{RB} configurations (i.e., signal bandwidths), considering a mismatched PA (i.e., voltage standing wave ratio, VSWR≈3). In order to cope with this challenging dynamic bandwidth re-allocation scenario we propose a IBW DPD model. Unlike other dynamic sizing DPD models (e.g., [3]) where all the coefficients associated to the new appended basis need to be re-trained and adjusted, in the proposed IBW generalized memory polynomial (GMP) model, the basis kernels are activated or deactivated without the need of

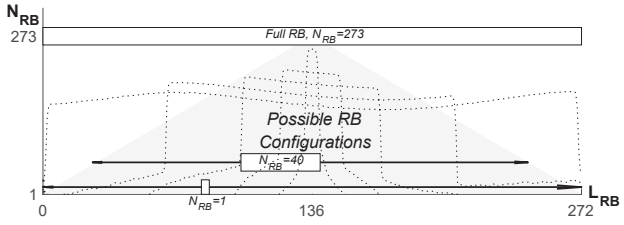

Fig. 1. Possible RB configurations for single band signals.

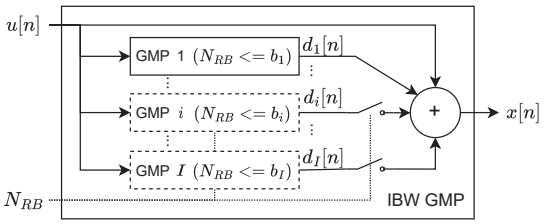

Fig. 2. Block diagram of IBW GMP model.

recalculating the coefficients values. Therefore, the coefficients values are fixed for the previous configurations, and new basis are only activated for new (higher) bandwidth configurations.

Therefore, the implementation of the IBW GMP model is straightforward by adding the functionality to activate or deactivate basis according to the signal bandwidth. First, we define the bandwidth boundary set $B = \{b_1, b_2, \cdots, b_i, \cdots, b_I\}$ in ascending order (i.e, $b_i > b_{i-1}$) to distinct the configurations for different bandwidths; where I is the total number of configurations and b_i is the maximum bandwidth (i.e., N_{RB}) for the i^{th} configuration. The additive distortion at the i^{th} configuration, $d_i[n]$, is modeled by the GMP as follows,

$$d_i\left[n\right] = \sum_{k_i=1}^{K_i} w_{k_i} u\left[n-\tau_{k_i}\right] \left|u\left[n-\tau_{k_i}-\sigma_{k_i}\right]\right|^{p_{k_i}} \quad (1)$$

where $u[n]$ is the discrete baseband input, τ_{k_i} and σ_{k_i} are the most significant baseband and lagging delays, respectively, p^{k_i} are the polynomial orders, and w_{k_i} are the corresponding coefficients. With this notation, for the i^{th} configuration, K_i basis are appended to the DPD model. For the full bandwidth configuration, the predistorted signal $x[n]$ is characterized by

$$x\left[n\right] = u\left[n\right] - \sum_{i=1}^{I} d_i\left[n\right]. \quad (2)$$

Fig. 2 shows the block diagram of the IBW GMP model that can enable or disable the corresponding GMP basis blocks according to the N_{RB} configuration of the incoming signal.

For extracting the DPD coefficients value (to be later used in open-loop), the training process begins with the lowest bandwidth configuration. The DPD coefficients are estimated to fit the residual distortion,

$$e_1[n] = u[n] - y[n] \quad (3)$$

where $u[n]$ and $y[n]$ are the baseband PA input and normalized (by the PA linear gain) output with signal bandwidth $N_{\mathrm{RB}} \leq b_1$. For higher bandwidth configurations, the coefficients at lower bandwidth configurations are fixed, and the estimation targets the residual error after employing those coefficients

$$e_i[n] = u[n] - y[n] - \sum_{j=1}^{i-1} d_j\left[n\right];\ (i>1). \quad (4)$$

The coefficients $\boldsymbol{w}_i = (w_{1_i} w_{2_i} \cdots w_{K_i})^T$ can be extracted following a closed-loop iterative least squares (LS) solution,

$$\boldsymbol{w}_i^{k+1} = \boldsymbol{w}_i^k + \mu(\boldsymbol{U}_i^H \boldsymbol{U}_i)^{-1} \boldsymbol{U}_i^H \boldsymbol{e}_i \quad (5)$$

Algorithm 1 IBW Constrained DOMP

procedure IBWDOMP($\boldsymbol{M}, \boldsymbol{U}, \boldsymbol{y}$)
  $\boldsymbol{M}^* \leftarrow \{\}, i \leftarrow 1, \boldsymbol{e} \leftarrow \boldsymbol{y}, \boldsymbol{w} \leftarrow ()$
  while $i \leq I$ **do**
    $\boldsymbol{M}_i \leftarrow \{\}, \boldsymbol{e}_i \leftarrow \boldsymbol{e}_{\{i\}}, \boldsymbol{Z}_i \leftarrow \boldsymbol{U}_{\{i\}}$
    repeat
      $\boldsymbol{\varphi}^{\{j\}} \xleftarrow{\forall j} \left| \frac{\boldsymbol{Z}_i{}^{\{j\}}}{\left\|\boldsymbol{Z}_i{}^{\{j\}}\right\|_2} \boldsymbol{e}_i \right|$
      $j^* \leftarrow \mathrm{pursuit}(\boldsymbol{M}^*, \boldsymbol{M}_i, \boldsymbol{M}, \boldsymbol{\varphi})$
      $\boldsymbol{M}_i \leftarrow \boldsymbol{M}_i \cup \boldsymbol{M}^{\{j^*\}}$
      $\boldsymbol{\sigma} \leftarrow \boldsymbol{Z}_i{}^H \boldsymbol{Z}_i$
      $\boldsymbol{Z}_i \leftarrow \boldsymbol{Z}_i - \boldsymbol{\sigma} \otimes \boldsymbol{Z}_i{}^{\{j^*\}}$
      $\boldsymbol{U}_* \leftarrow \boldsymbol{U}_i\left[\boldsymbol{M}_i\right]$
      $\boldsymbol{w}_i \leftarrow \left(\boldsymbol{U}_*^{\boldsymbol{H}} \boldsymbol{U}_*\right)^{-1} \boldsymbol{U}_*^{\boldsymbol{H}} \boldsymbol{y}_i$
      $\boldsymbol{e}_i \leftarrow \boldsymbol{y}_i - \boldsymbol{U}_* \boldsymbol{w}_i$
    until stopping criterion is met
    $\boldsymbol{M}^* \leftarrow \boldsymbol{M}^* \cup \boldsymbol{M}_i$
    $\boldsymbol{U}_* \leftarrow \boldsymbol{U}\left[\boldsymbol{M}^*\right]$
    $\boldsymbol{w} \leftarrow \left(\boldsymbol{w}^T, \boldsymbol{w}_i{}^T\right)^T$
    $\boldsymbol{e} \leftarrow \boldsymbol{y} - \boldsymbol{U}_* \boldsymbol{w}$
    $\boldsymbol{M} \leftarrow \boldsymbol{M}_{\notin \boldsymbol{M}_i}$ ▷ only if not allowing re-selection
    $i \leftarrow i + 1$
  end while
  return $\boldsymbol{M}^*$
end procedure

where $(\cdot)^H$ denotes Hermitian transpose, $\boldsymbol{U}_i = (\boldsymbol{\psi}_i\left[0\right], \cdots, \boldsymbol{\psi}_i\left[n\right], \cdots, \boldsymbol{\psi}_i\left[N-1\right])^T$, is the $N \times K_i$ data matrix, with $n = 0, \cdots, N-1$. The vector containing the basis functions in (1) is $\boldsymbol{\psi}_i[n] = \left(\psi_{i_1}\left[n\right], \cdots, \psi_{i_j}\left[n\right], \cdots, \psi_{i_{K_i}}\left[n\right]\right)$, while $\mu \leq 1$ is the learning rate and k the iteration number.

Ideally, we would like to have independent (i.e., orthogonal) basis to capture the remaining linear and non-linear distortion at every bandwidth configuration. However, this is not possible because modeling memory effects (i.e., frequency response) requires the use of highly correlated basis. Therefore, the objective is to find the optimal DPD basis functions orthogonal to the residual error at each bandwidth configuration, i.e., finding the optimal delays and polynomial orders in (1) for each IBW configuration. For this purpose, this paper proposes a constrained DOMP algorithm for IBW GMP basis selection, as described in the following section.

III. Constrained DOMP for Basis Selection

In this paper, we extended the DOMP algorithm in [2] to constrain the basis selection in a dynamic bandwidth re-allocation scenario. First, we define $\boldsymbol{M}^* = \{\boldsymbol{M}_1 \cup \cdots \cup \boldsymbol{M}_i \cup \cdots \cup \boldsymbol{M}_I\}$ as the optimal kernel set storing the basis delays and polynomial orders for each IBW GMP model, where $\boldsymbol{M}_i = \{m_{1_i} \cup \cdots \cup m_{k_i} \cup \cdots \cup m_{K_i}\}$, with K_i being the number of basis for the i^{th} configuration. Each element in the kernel set $m_{k_i} = (\tau_{k_i}, \sigma_{k_i}, p_{k_i})$ stores the baseband delay, lagging delay and the polynomial order of the basis, in consistency with (1). The constrained DOMP process for IBW GMP is described in Algorithm 1. The input $\boldsymbol{M}$ is the kernel set containing all the basis candidates; while $\boldsymbol{U} = (\boldsymbol{U}_1{}^T, \cdots, \boldsymbol{U}_i{}^T, \cdots \boldsymbol{U}_I{}^T)^T$ is the data matrix build according

Algorithm 2 Greedy selection with memory preference

1: **procedure** PURSUIT($\boldsymbol{M}^*, \boldsymbol{M}_i, \boldsymbol{M}, \boldsymbol{\varphi}$)
2: $\quad \boldsymbol{P}^* \leftarrow \boldsymbol{M}^{*\{\tau,\sigma\}} \cup \boldsymbol{M}_i{}^{\{\tau,\sigma\}}, \boldsymbol{P} \leftarrow \boldsymbol{M}^{\{\tau,\sigma\}}$
3: $\quad \boldsymbol{\varphi}_n \leftarrow \frac{\boldsymbol{\varphi}-\min(\boldsymbol{\varphi})}{\max(\boldsymbol{\varphi}-\min(\boldsymbol{\varphi}))}$
4: $\quad \boldsymbol{\varphi}_s, \boldsymbol{s}_{\mathrm{id}} \xleftarrow{\forall \boldsymbol{\varphi}_n > \gamma} \mathrm{sort}_{\mathrm{desc}}(\boldsymbol{\varphi}_n)$
5: $\quad \boldsymbol{P}_s \leftarrow \boldsymbol{P}^{\{\boldsymbol{s}_{id}\}}$
6: $\quad$ **if** $\boldsymbol{P}_s \cap \boldsymbol{P}^* \neq \emptyset$ **then**
7: $\quad\quad j \leftarrow 0$
8: $\quad\quad$ **repeat**
9: $\quad\quad\quad j \leftarrow j + 1$
10: $\quad\quad\quad j^* \leftarrow \boldsymbol{s}_{id}(j)$
11: $\quad\quad$ **until** $\boldsymbol{P}_s^{\{j\}} \in \boldsymbol{P}^*$
12: $\quad$ **else**
13: $\quad\quad j^* \leftarrow \boldsymbol{s}_{id}(1)$
14: $\quad$ **end if**
15: $\quad$ **return** j^*
16: **end procedure**

to $\boldsymbol{M}$, stacking by rows the data matrices of all different bandwidth configurations; the vector of the PA outputs is defined as $\boldsymbol{y} = \left(\boldsymbol{y}_1^T, \cdots, \boldsymbol{y}_i^T, \cdots, \boldsymbol{y}_I^T\right)^T$. The operation $(\cdot)_{\{i\}}$ selects rows corresponding to the i^{th} bandwidth configuration. For example, $\boldsymbol{y}_i = \boldsymbol{y}_{\{i\}}$ selects from $\boldsymbol{y}$ the outputs with signal bandwidths $b \in (b_{i-1}, b_i]$. The operation $(\cdot)^{\{j\}}$ selects the j^{th} column or element from the data matrix or data set. In line 11 of Algorithm 1, the $\boldsymbol{U}_i[\boldsymbol{M}_i]$ selects columns from $\boldsymbol{U}_i$, in particular the ones indexed by the kernel set $\boldsymbol{M}_i$; and the same operation applies to line 16. Line 10 performs the Gram-Schmidt process, where $\otimes$ denotes the Kronecker product. Line 19 removes the selected basis from the candidates set if we do not allow basis being re-selected in the higher bandwidth configuration. This only becomes true if the additive distortions follow the orthogonality assumption.

In Algorithm 1, the pursuit in line 7 and the stopping criterion in line 14 are included and particularized in this paper, taking into account the implementation requirements and constraints in handsets. The original DOMP selects the most relevant non-repeated basis (i.e., the one with the highest score in $\boldsymbol{\varphi}$) from the candidates set,

$$j^* = \arg \max_{\boldsymbol{M}^{\{j\}} \notin \boldsymbol{M}_i} \boldsymbol{\varphi} \tag{6}$$

where j^* is the index of the recommended basis. Since there is no constraint on the memory delays combination, in this incremental bandwidth application, it tends to select a lot of memory pairs (i.e., unique combinations of $(\tau_{k_i}, \sigma_{k_i})$ regardless of the polynomial order). From the implementation point of view, assuming the direct implementation on FPGA [4], it complicates the routing of programmable logic (PL) and makes it difficult to meet the timing. In addition, the score variation with memory in $\boldsymbol{\varphi}$ is not significant when operating with a narrow-band signal due to the oversampling. Instead, the polynomial order dominates over the memory delays.

Consequently, we introduced memory preference in the pursuit, as shown in Algorithm 2. In line 2, by discarding the polynomial orders, delay pairs set $\boldsymbol{P}^*$ and $\boldsymbol{P}$ are extracted from the selected basis set and the candidate basis set. The score vector $\boldsymbol{\varphi}$ is normalized to 1 with offset removed in line 3. Then, in line 4, the normalized scores are sorted in descending order and those high scores above the threshold γ are selected. The threshold was set to $\gamma = 0.8$ in this work. If we can find repeated memory pairs from these high scores memory pairs ($\boldsymbol{P}_s$) in the already selected memory pairs ($\boldsymbol{P}^*$), the first repeated one will be selected. Otherwise, the first one with the highest score will be selected, and a new memory pair is added. This selection pursuit can effectively reduce the total number of memory pairs after the DOMP process.

For the stopping criterion, instead of using the Bayesian information criterion, which usually suggest a lot of basis, a modeling NMSE increment criterion is proposed. It works as follows. If by adding a new basis, the improvement of the modeling NMSE is lower than a threshold ρ, the stopping criterion is met. In addition, the stopping criterion also specifies the maximum number of basis allowed for each IBW configuration. This is defined according to the expected available resources for the hardware implementation.

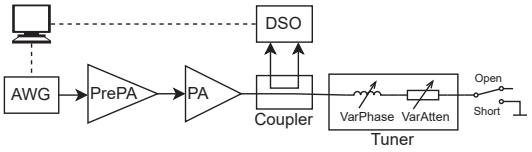

Fig. 3. Block diagram of the experimental testbed.

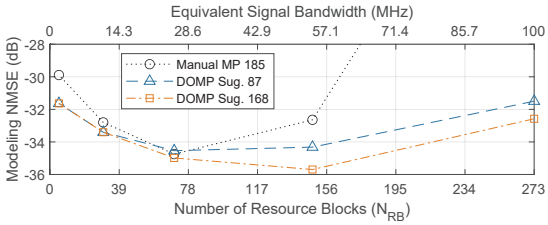

Fig. 4. Modeling performance of the IBW DPD models.

IV. Experimental Setup and Results

Fig. 3 shows the diagram Matlab-controlled PA testbench, where the arbitrary waveform generator M8190 and the DSO90404A oscilloscope from Keysight were used for waveform generation and acquisition, respectively. The driver was a ZHL-42 from Mini-Circuits and the PA a SoC developed by Hisilicon for handset applications. The non-flat response was achieved forcing a mismatched condition (VSWR ≈ 3), by using a tuner composed of a phase tuner and a variable attenuator with an open end. The PA was operated with a NR QPSK OFDM signal with 30 KHz subcarrier spacing and 11 dB of PAPR. The bandwidth configurations for training is $N_{\mathrm{RB,train}} = \{5, 30, 70, 148, 273\}$, and the IBW bandwidth boundary $B = \{10, 40, 80, 162, 273\}$.

Fig. 4 shows the modeling performance comparing the basis suggested by DOMP (allowing reselection) and the manually constructed IBW MP model. The manual model appends nearby memories when the signal bandwidth increases but, as observed, it fails to model the full N_{RB} signal. The 3D stem plot in Fig. 5 shows how the selected basis are appended

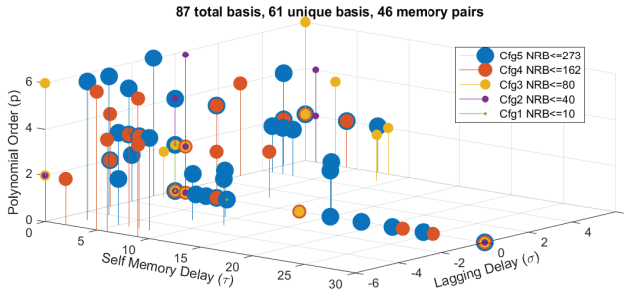

Fig. 5. The IBW basis appended at each bandwidth configuration.

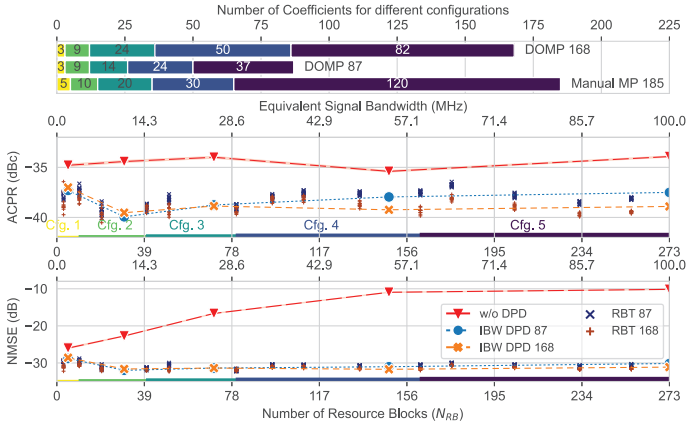

Fig. 6. IBW DPD and robustness performance with IBW DPD models.

to the model at each bandwidth configuration (few of them are re-selected). By including the memory preference in the pursuit, only 46 memory pairs are selected while, without the memory preference constrain, 56 memory pairs are required.

Fig. 6-top shows the number of coefficients employed for each IBW configuration. As observed in Fig. 6-bottom, with the proposed constrained DOMP algorithm we can trade-off the IBW DPD computational complexity (in terms of number of coefficients) and the linearization performance (i.e., ACPR and NMSE), being more critical when considering the full bandwidth (100 MHz or $N_{\mathrm{RB}} = 273$) signal. To guarantee open-loop operation, we conducted a robustness test (RBT) considering 10 different signals with different bandwidth configurations that were not used for training the IBW DPD model. As observed, for all test cases it was possible to meet the linearity specification of ACPR<-36 dBc.

V. Conclusion

An open-loop IBW DPD architecture for dynamic bandwidth scenarios was presented in this paper. A constrained DOMP algorithm was proposed for IBW basis selection. Targeting an efficient hardware implementation, the scalable IBW GMP model only uses very few basis for narrow band signals, while for wider or full bandwidth (100 MHz) signals it can activate further or all basis functions. The scalable DPD approach was validated with a handset PA SoC presenting a strong non-flat response (load-mismatched) condition. Robustness tests considering different signal bandwidths prove the open-loop consistency of the proposed IBW DPD.

Acknowledgment

The work in this paper was funded by Huawei Technologies from January 2022 to December 2022; and supported in part by the project PID2020-113832RB-C21 funded by MCIN/AEI/10.13039/50110001103 and in part by the Government of Catalonia and the European Social Fund under Grant 2021-FI-B-137.

References

[1] X. Liu, W. Chen, W. Chen, Y. Guo, and Z. Feng, "Load-mismatch tracking digital predistortion for mobile-terminal power amplifiers," in *in Proc. IEEE MTT-S Int. Microw. Symp. Dig. (IMS)*, Jun. 2022.

[2] J. A. Becerra, M. J. Madero-Ayora, J. Reina-Tosina, C. Crespo-Cadenas, J. Garcia-Frias, and G. Arce, "A doubly orthogonal matching pursuit algorithm for sparse predistortion of power amplifiers," *IEEE Microw. Wireless Compon. Lett.*, vol. 28, no. 8, pp. 726–728, 2018.

[3] Y. Li and A. Zhu, "On-demand real-time optimizable dynamic model sizing for digital predistortion of broadband RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 7, pp. 2891–2901, 2020.

[4] W. Li, N. Bartzoudis, J. R. Fernández, D. López-Bueno, G. Montoro, and P. L. Gilabert, "FPGA implementation of a linearization system for wideband envelope tracking power amplifiers," *IEEE Trans. Microw. Theory Techn.*, pp. 1–12, 2022.