A Low-Complexity Post-Weighting Predistorter in a mMIMO Transmitter Under Crosstalk

Ganesh Prasad, *Member, IEEE* and Håkan Johansson, *Senior Member, IEEE*

Abstract— The beam-oriented digital predistortion (BO-DPD) is not sufficient to linearize the output from a subarray of power amplifiers (PAs) in different directions except the desired direction. Therefore, subsequent to the BO-DPD operation, we perform a post-weighting (PW) processing to minimize the nonlinear radiations in the wide range of directions under crosstalk. Here, the optimized PW coefficients are multiplied by the polynomial terms of the BO-DPD, then, the resultant signals are distributed to the PAs to compensate the nonlinear radiations. In this work, first, we propose fully-featured post-weighting (FF-PW) scheme, then, we derive a low-complexity post-weighting (LC-PW) scheme.

Index Terms—Predistortion, polynomial model, subarray of PAs, post-weighting, convex optimization

Fig. 1. An architecture of a DPD followed by a PW predistortion.

I. INTRODUCTION

For efficient transmission of signals, radio frequency (RF) power amplifiers (PAs) play an important role. However, the design of highly linear PAs over the large dynamic range of the signals is expensive. Further, it is costly for a massive multiple-input multiple-output (mMIMO) transmitter with a large number of PAs under crosstalk. So, the linearization requires a less complex and proficient predistortion scheme.

Initially, the multiple antennas transmitter was linearized using a single digital predistortion (DPD) by considering equal nonlinear characteristics of its PAs [1]. However, in practice, the nonlinear characteristics are not equal and later, instead of fully linearizing all PAs, the beam-oriented (BO) output in a desired direction was linearized using a single DPD, known as BO-DPD[1] under crosstalk [2], [3]. But, it is not able to provide the linearization in other directions except the desired direction, thus, it gives nonlinear sidelobes in the BO output. Then, it is realized that a single DPD with one predistorted output signal is not sufficient to linearize all the PAs of different nonlinearites [4]. Later, the full linearization of the PAs was achieved by including a tuning box to each PA. But, it has high complexity as each tuning box requires a training [5].

Recently, in a post-weighting (PW) scheme, a single DPD training followed by a post-weighting (PW) optimization is used to generate more than one predistorted signal in a subarray, then, these signals are distributed to multiple PAs to address their nonlinearities [6]. But, it provides only one PW coefficient (one degree of freedom (DOF)) per PA that is not sufficient to linearize the multiple PAs. In the proposed PW scheme, the DOF per PA is increased with less complexity

G. Prasad and H. Johansson are with the Division of Communication Systems, Department of Electrical Engineering, Linköping University, 581 83 Linköping, Sweden (e-mail: {ganesh.prasad, hakan.johansson}@liu.se).

[1]In this work, the terms DPD and BO-DPD are used interchangeably.

by reducing the adders, multipliers, and the RF chains. In this regard, the key contribution of this work is three-fold. (i) First, using the dual-input polynomial models of the PAs, an approximate relationship is established between the crosstalks at the PAs and the crosstalk compensation signal to the BO-DPD. Then, using it, we train the DPD. (ii) Next, we propose a fully-featured PW (FF-PW) scheme and using it, we derive a low-complexity PW (LC-PW) scheme. Based on it, the system parameters are arranged non-trivially into suitable vectors and matrices to simplify the system analysis. (iii) Further, an expression for nonlinear radiation from the BO transmitter operating with DPD and PW is obtained. Using it, a convex minimization problem is formulated. Then, its optimal PW coefficients are obtained in a closed form. Finally, numerical results are obtained to get various design insights.

II. SYSTEM DESCRIPTION AND BO-DPD TRAINING

A. System Architecture

In a mMIMO transmitter, we consider a uniform linear array of $K \times S$ PAs where each of K subarrays contains S PAs. The message vector that needs to be transmitted is $\boldsymbol{s} = [s_1, \cdots, s_K]^T$, where s_k is the message to be transmitted by the kth subarray. Fig. 1 represents an architecture for the linearization of the kth subarray under crosstalk signals $\{c_{kl}\}$; $l \in \{1, \cdots, S\}$ using two layers of predistortion operations: DPD followed by the PW predistortion. Here, c_{kl} is the crosstalk signal at lth PA of the kth subarray due to transmit signals from the remaining PAs. In this scheme, first, the DPD is trained separately for the BO output from the subarray. Thus, during the training of the DPD, the PAs are directly connected to the DPD and isolated from the PW block by closing the switch $\mathcal{S}_0$ and opening the switches, $\{\mathcal{S}_i\}$; $i \in \{1, \cdots, N_{RF}\}$, where N_{RF} $(\leq S)$ is the number of RF chains in the subarray. In the training, the message s_k and

the signal c_k are inputted to the dual-input DPD, where c_k is obtained from the training block as shown in the figure. Using them, the DPD generates the predistorted signal x_k to address the nonlinearties of the PAs under the crosstalks $\{c_{kl}\}$. Using Fig. 1, the generation of x_k can be described as follows. The dual-input DPD is modeled using a polynomial model (as described later using (1a)), having $Q+1$ terms which are its outputs. Each term is a basis function (of s_k and c_k) multiplied by its coefficient. The $Q+1$ coefficients are represented by a vector $\mathbf{\Phi}_k$ which needs to be trained in the training block. The resultant predistorted signal from the DPD is x_k which is the sum of $Q+1$ outputs (cf. (1a) in Section II-B)[2]. Thereafter, x_k is multiplied by the analog beamforming weights, $\{w_{kl}\}$ to get the BO output in a given direction, where w_{kl} ($|w_{kl}| = 1$) is the analog phase shifter to the lth PA of the kth subarray and they are represented in a vector $\mathbf{w}_k = [w_{k1}, \cdots, w_{kS}]^T$. The BO output is feedback to the training block to train the coefficients $\mathbf{\Phi}_k$ and the signal c_k using the iterative method as described in Section II-C. At the end of the training, their estimated values, $\widehat{\mathbf{\Phi}}_k$ and $\widehat{c}_k$ are obtained. Using them, the trained DPD provides the $Q+1$ outputs that are connected to the PAs via the PW block after opening $\mathcal{S}_0$ and closing $\{\mathcal{S}_i\}$. In the PW block, the PW coefficients (dedicated for each $n_{PA} = S/N_{RF}$ PAs to enhance the predistortion DOF) are multiplied by the Q nonlinear outputs of the DPD. Thereafter, the PW block outputs the N_{RF} number of signals and each of these signals are distributed to n_{PA} number of PAs to get the desired BO output.

B. Dual Input Polynomial Model of DPD and PAs

We consider the dual-input memoryless polynomial models[3] [2] for the the DPD and PAs. After omitting the nonlinear terms of the signal c_k, the output x_k of the DPD as a function of s_k and c_k can be expressed as in (1a) which is further represented in matrix form in (1b).

$$x_k = \sum_{p=0}^{(P-1)/2}\phi_{kp}^0\psi_p^0(s_k) + \sum_{p=0}^{(P-1)/2}\phi_{kp}^1\psi_p^1(s_k)c_k + \sum_{p=1}^{(P-1)/2}\phi_{kp}^2\psi_p^2(s_k)c_k^*, \quad(1a)$$

$$x_k = \breve{\mathbf{\Psi}}(s_k, c_k)\breve{\mathbf{\Phi}}_k; \quad k \in \{1, 2, \cdots, K\}, \quad(1b)$$

where P is the order of the polynomial[4] and $\breve{\mathbf{\Psi}}(s_k, c_k) = [\mathbf{\Psi}^0(s_k), \mathbf{\Psi}^1(s_k)c_k, \mathbf{\Psi}^2(s_k)c_k^*]$ is the row vector of basis functions, $\mathbf{\Psi}^v(s_k) = [\psi_\mu^v(s_k), \cdots, \psi_{(P-1)/2}^v(s_k)]$ for $v \in \{0, 1, 2\}$. $\psi_p^0(s_k) = s_k^{p+1}s_k^{*p}$, $\psi_p^1(s_k) = s_k^p s_k^{*p}$, and $\psi_p^2(s_k) = s_k^{p+1}s_k^{*p-1}$. μ is the initial value of p and $\mu = 1$ for $v = 2$; otherwise, $\mu = 0$. $\breve{\mathbf{\Phi}}_k$ is a column vector of the coefficients for the basis functions in $\mathbf{\Psi}(s_k, c_k)$, given by $\breve{\mathbf{\Phi}}_k = [\mathbf{\Phi}_k^{0T}, \mathbf{\Phi}_k^{1T}, \mathbf{\Phi}_k^{2T}]^T$. $\mathbf{\Phi}_k^v = [\phi_{k\mu}^v, \cdots, \phi_{k(P-1)/2}^v]^T$ and ϕ_{kp}^v is the coefficient of basis $\psi_p^v(s_k)$. Note that $\breve{\mathbf{\Psi}}(s_k, c_k)$ and $\breve{\mathbf{\Phi}}_k$ contain the basis functions and coefficients for the polynomial of order P in (1a). However, in practice, certain basis

---

[2]A conventional DPD has one output x_k based on a polynomial model [2]. But, here, each term of the polynomial is further processed by the PW block (cf. Fig. 1), thus, the DPD has $Q+1$ outputs and x_k is the sum of them.

[3]For simplicity, we consider the memoryless polynomial models, however, the proposed work is equally applicable for memory polynomial models.

[4]Although, for simplicity, the orders of the polynomials are represented by the same symbol, P in (1a) and (2a), they can take different values.

functions with their nonzero coefficients play the dominant role in predistortion for a given type of PAs. Thus, in general, if $(Q+1)$ basis functions have their dominant role in the predistortion, then, the basis row vector $\mathbf{\Psi}(s_k, c_k)$ is defined as: $\mathbf{\Psi}(s_k, c_k) \triangleq [\psi_{p_1}^{v_1}\mathcal{C}_k^{v_1}, \cdots, \psi_{p_{(Q+1)}}^{v_{(Q+1)}}\mathcal{C}_k^{v_{(Q+1)}}]$ and corresponding coefficient column vector is: $\mathbf{\Phi}_k \triangleq [\phi_{kp_1}^{v_1}, \cdots, \phi_{kp_{(Q+1)}}^{v_{(Q+1)}}]^T$. Here, $\mathcal{C}_k^{v_i} = \delta(v_i) + c_k\delta(v_i - 1) + c_k^*\delta(v_i - 2)$, $p_i \in \{0, \cdots, (P-1)/2\}$, $v_i \in \{0, 1, 2\}$ for $i \in \{1, \cdots, (Q+1)\}$, Q is the number of nonlinear basis functions, and only one basis function, $\psi_{p_i}^{v_i}\mathcal{C}_k^{v_i} = s_k$ is linear for $p_i = v_i = 0$ (thus, the total number of basis functions is $(Q+1)$). $\delta(\cdot)$ is the Kronecker delta function. Hereafter, we consider the DPD in $\mathbf{\Psi}(s_k, c_k)$ and $\mathbf{\Phi}_k$. Moreover, output signal vector of K DPDs is denoted as $\boldsymbol{x} = [x_1, \cdots, x_K]^T$.

Similarly, based on dual input memoryless polynomial model, for the inputs, x_k and c_{kl} to the lth PA of the kth subarray, its output y_{kl} is expressed as in (2a) and its matrix form in (2b).

$$y_{kl} = \sum_{p=0}^{(P-1)/2}\phi_{klp}^0\psi_p^0(w_{kl}x_k) + \sum_{p=0}^{(P-1)/2}\phi_{klp}^1\psi_p^1(w_{kl}x_k)c_{kl}$$
$$+ \sum_{p=1}^{(P-1)/2}\phi_{klp}^2\psi_p^2(w_{kl}x_k)c_{kl}^* \quad(2a)$$

$$y_{kl} = \breve{\mathbf{\Psi}}(w_{kl}x_k, c_{kl})\breve{\mathbf{\Phi}}_{kl}; \quad l \in \{1, 2, \cdots, S\}, \quad(2b)$$

where $\breve{\mathbf{\Psi}}(w_{kl}x_k, c_{kl})$ and $\breve{\mathbf{\Phi}}_{kl}$ are the row and column vectors of the basis functions and its coefficients respectively. They can be represented similarly as described above for the DPD basis functions and coefficients with additional suffix l to represent it for lth PA of the subarray. Further, for a given type of PAs, similar to the DPD, the $Q'+1$ dominant basis functions to identify them can be represented in a row vector $\mathbf{\Psi}(w_{kl}x_k, c_{kl})$ and corresponding coefficients in a column vector can be given by $\mathbf{\Phi}_{kl}$. Note that the PAs coefficients $\mathbf{\Phi}_{kl}$; $l \in \{1, \cdots, S\}$ are assumed to be known which can be identified using least square (LS) estimation as described briefly in the numerical section. Further, all the outputs from the PAs can be expressed in a matrix form as:

$$\mathbf{Y}_k = \mathbf{\Omega}_k\mathbf{\Theta}_k, \quad(3)$$

where $\mathbf{Y}_k = [y_{k1}, \cdots, y_{kS}]^T$, $\mathbf{\Omega}_k = \text{diag}([\mathbf{\Psi}(w_{k1}x_k, c_{k1})^T, \cdots, \mathbf{\Psi}(w_{kS}x_k, c_{kS})^T]^T)$ and $\mathbf{\Theta}_k = [\mathbf{\Phi}_{k1}^T, \cdots, \mathbf{\Phi}_{kS}^T]^T$.

Using (3), for the steering vector $\mathbf{h}_k^\varphi$ at angle (direction) φ to the vertical plane of the array, the BO signal z_k^φ is:

$$z_k^\varphi = \mathbf{h}_k^{\varphi T}\mathbf{Y}_k = \mathbf{h}_k^{\varphi T}\mathbf{\Omega}_k\mathbf{\Theta}_k, \quad(4)$$

where $\mathbf{h}_k^\varphi = [h_{k1}^\varphi, \cdots, h_{kS}^\varphi]^T$ and h_{kl}^φ is the lth steering element. As c_{kl} cannot be measured at the lth PA, it can be expressed as a linear combination of the transmit signals from the other PAs as [2]:

$$c_{kl} = \sum_{i=1}^K\sum_{r=1}^S \lambda_{kl,ir}' y_{ir} = \boldsymbol{\lambda}_{kl}'\mathbf{Y}, \quad(5)$$

where $\boldsymbol{\lambda}_{kl}' = [\boldsymbol{\lambda}_{kl,1}', \cdots, \boldsymbol{\lambda}_{kl,K}']$, and $\boldsymbol{\lambda}_{kl,i}' = [\lambda_{kl,i1}', \cdots, \lambda_{kl,iS}']$, and $\mathbf{Y} = [\mathbf{Y}_1^T, \cdots, \mathbf{Y}_K^T]^T$. $\lambda_{kl,ir}'$ is the coefficient for contribution in crosstalk signal c_{kl} at lth PA of the kth subarray from rth PA of the ith subarray. Simplifying (5) in linear terms of c_{ir} and x_k after omitting the negligible nonlinear terms, we get c_{kl} as in (6a). Further, it is expressed in a matrix form in (6b), then, simplified

to (6c).

$$c_{kl} = \sum_{i=1}^{K} \sum_{r=1}^{S} \left[\lambda'_{kl,ir} \phi^0_{ir0} w_{ir} x_i + \lambda'_{kl,ir} \phi^1_{ir0} c_{ir} \right] \quad (6a)$$

$$\boldsymbol{c} = \boldsymbol{A}^0 W_D \boldsymbol{x} + \boldsymbol{A}^1 \boldsymbol{c} \quad (6b)$$

$$\Rightarrow \boldsymbol{c} = (I - \boldsymbol{A}^1)^{-1} \boldsymbol{A}^0 W_D \boldsymbol{x} = \boldsymbol{\Lambda} W_D \boldsymbol{x}, \quad (6c)$$

where $\boldsymbol{c} = [\overline{\boldsymbol{c}}_1^T, \cdots, \overline{\boldsymbol{c}}_K^T]^T$, $\overline{\boldsymbol{c}}_k = [c_{k1}, \cdots, c_{kS}]^T$, $\boldsymbol{A}^v = \overline{\boldsymbol{\Lambda}} \mathrm{diag}(\overline{\boldsymbol{\Phi}}_1^v, \cdots, \overline{\boldsymbol{\Phi}}_K^v)$, $\overline{\boldsymbol{\Phi}}_k^v = [\phi^v_{k10}, \cdots, \phi^v_{kS0}]$ for $v \in \{0,1\}$, $\overline{\boldsymbol{\Lambda}} = [\boldsymbol{\lambda}'^T_{11}, \cdots, \boldsymbol{\lambda}'^T_{1S}, \cdots, \boldsymbol{\lambda}'^T_{K1}, \cdots, \boldsymbol{\lambda}'^T_{KS}]^T$, and $W_D = \mathrm{diag}([w_{11}, \cdots, w_{1S}]^T, \cdots, [w_{K1}, \cdots, w_{KS}]^T)$. From (6c), $\boldsymbol{\Lambda} = (I - \boldsymbol{A}^1)^{-1} \boldsymbol{A}^0$ which is nothing but the coefficients associated with the weighted signal vector $W_D \boldsymbol{x}$ to get $\boldsymbol{c}$. It can be expressed as $\boldsymbol{\Lambda} = [\boldsymbol{\Lambda}_1^T, \cdots, \boldsymbol{\Lambda}_K^T]^T$, where $\boldsymbol{\Lambda}_k = [\boldsymbol{\lambda}_{k1}^T, \cdots, \boldsymbol{\lambda}_{kS}^T]^T$ and $\boldsymbol{\lambda}_{kl} = [\lambda_{kl,11}, \cdots, \lambda_{kl,1S}, \cdots, \lambda_{kl,K1}, \cdots, \lambda_{kl,KS}]$. Next, using (4) and (6c), we describe the training of the BO-DPD.

C. Training of BO-DPD

In [2], the relationship between c_k and $\overline{c}_k$ is obtained without considering $\boldsymbol{w}_k$. To include $\boldsymbol{w}_k$, first, we consider the same approximation, $\lambda_{kl,ir} \approx \alpha_l \lambda_{ki}$ due to uniform and linear arrangement of PAs [2]. It denotes that the crosstalk from rth PA of the ith subarray to lth PA of the kth subarray with coefficient $\lambda_{kl,ir}$ can be approximated to α_l times the overall crosstalk from the ith subarray to kth subarray with coefficient λ_{ki}. Applying it to $\boldsymbol{\lambda}_{kl}$ (cf. Section II-B), we get: $\boldsymbol{\lambda}_{kl} \approx \alpha_l [\underbrace{\lambda_{k1}, \cdots, \lambda_{k1}}_{S \text{ times}}, \cdots, \underbrace{\lambda_{kK}, \cdots, \lambda_{kK}}_{S \text{ times}}] = \alpha_l [\lambda_{k1} \mathbf{1}_S^T, \cdots, \lambda_{kK} \mathbf{1}_S^T] = \alpha_l \boldsymbol{\lambda}_k^T D_{\mathbf{1}}$, where $\boldsymbol{\lambda}_k = [\lambda_{k1}, \cdots, \lambda_{kK}]^T$, $D_{\mathbf{1}} = \mathrm{diag}(\underbrace{\mathbf{1}_S^T, \cdots, \mathbf{1}_S^T}_{K \text{ times}})$ and $\mathbf{1}_S$ is the column vector of ones of length S. Thus, $\boldsymbol{\Lambda}_k$ (cf. Section II-B) is: $\boldsymbol{\Lambda}_k \approx \boldsymbol{\alpha} \boldsymbol{\lambda}_k^T D_{\mathbf{1}}$, where $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_S]^T$. After applying this approximation in (6c), we get:

$$\overline{\boldsymbol{c}}_k = \boldsymbol{\alpha} c_k; \quad c_k = \boldsymbol{x}^T W_D^T D_{\mathbf{1}}^T \boldsymbol{\lambda}_k. \quad (7)$$

Using (7), the inputs to the basis functions in $\boldsymbol{\Omega}_k$ of (3) can be expressed in x_k and c_k. As c_k depends on $\boldsymbol{\lambda}_k$, we estimate $\boldsymbol{\lambda}_k$ to find c_k. In this regard, using (7), z_k^φ in (4) is given as:

$$z_k^\varphi = g_k^0 + \boldsymbol{g}_k^1 \widehat{\boldsymbol{\lambda}}_k + \boldsymbol{g}_k^2 \widehat{\boldsymbol{\lambda}}_k^*, \quad (8)$$

where $g_k^0 = \sum_{l=1}^{S} h_{kl}^\varphi \boldsymbol{\Psi}^0(w_{kl} x_k) \widehat{\boldsymbol{\Phi}}_{kl}^0$, $\boldsymbol{g}_k^1 = \sum_{l=1}^{S} h_{kl}^\varphi \boldsymbol{\Psi}^1(w_{kl} x_k) \widehat{\boldsymbol{\Phi}}_{kl}^1 \boldsymbol{x}^T W_D^T D_{\mathbf{1}}^T$, and $\boldsymbol{g}_k^2 = \sum_{l=1}^{S} h_{kl}^\varphi \boldsymbol{\Psi}^2(w_{kl} x_k) \widehat{\boldsymbol{\Phi}}_{kl}^2 \boldsymbol{x}^H W_D^H D_{\mathbf{1}}^H$. Further, by including time samples of $\boldsymbol{x}$, (8) is given by:

$$\boldsymbol{z}_k^\varphi = \boldsymbol{g}_k^0 + \boldsymbol{G}_k^1 \widehat{\boldsymbol{\lambda}}_k + \boldsymbol{G}_k^2 \widehat{\boldsymbol{\lambda}}_k^*, \quad (9)$$

where z_k^φ, g_k^0, $\boldsymbol{g}_k^1$, and $\boldsymbol{g}_k^2$ are denoted as $\boldsymbol{z}_k^\varphi$, $\boldsymbol{g}_k^0$, $\boldsymbol{G}_k^1$, and $\boldsymbol{G}_k^2$ respectively after including the time samples[5]. By splitting (9) into real and imaginary parts, the real part $\mathcal{R}(\widehat{\boldsymbol{\lambda}}_k)$ and the imaginary part $\mathcal{I}(\widehat{\boldsymbol{\lambda}}_k)$, can be determined as:

$$\begin{bmatrix} \mathcal{R}(\widehat{\boldsymbol{\lambda}}_k) \\ \mathcal{I}(\widehat{\boldsymbol{\lambda}}_k) \end{bmatrix} = \begin{bmatrix} \mathcal{R}(\boldsymbol{G}_k^1 + \boldsymbol{G}_k^2) & \mathcal{I}(-\boldsymbol{G}_k^1 + \boldsymbol{G}_k^2) \\ \mathcal{I}(\boldsymbol{G}_k^1 + \boldsymbol{G}_k^2) & \mathcal{R}(\boldsymbol{G}_k^1 - \boldsymbol{G}_k^2) \end{bmatrix}^\dagger \begin{bmatrix} \mathcal{R}(\boldsymbol{z}_k^\varphi - \boldsymbol{g}_k^0) \\ \mathcal{I}(\boldsymbol{z}_k^\varphi - \boldsymbol{g}_k^0) \end{bmatrix} \quad (10)$$

To determine $\widehat{\boldsymbol{\lambda}}_k$, first, $\boldsymbol{x}_k$ is set as: $\boldsymbol{x}_k = \boldsymbol{s}_k$. Thus, we can compute $\boldsymbol{G}_k^1$, $\boldsymbol{G}_k^2$, and $\boldsymbol{g}_k^0$ and using the measured $\boldsymbol{z}_k^\varphi$,

[5]Note that in this work, the boldface symbol $\boldsymbol{a}$ represents the scalar signal a with its time samples.

Fig. 2. Flow diagrams for the identification of DPD coefficients.

Fig. 3. Two PW schemes, (a) FF-PW and (b) LC-PW for $S = 4$.

from (10), we find $\widehat{\boldsymbol{\lambda}}_k$. Further, the obtained $\widehat{\boldsymbol{\lambda}}_k$ is used in the algorithm as shown in Fig. 2. Here, initially, we again set $\boldsymbol{x}_k = \boldsymbol{s}_k$ and for given $\widehat{\boldsymbol{\lambda}}_k$, c_k is determined. Then, using post-inverse, the measured $\boldsymbol{z}_k^\varphi$ and c_k are set as inputs to the DPD to find $\boldsymbol{\Phi}_k$ using LS method. Using it, the output $\boldsymbol{x}_k$ of the DPD and c_k are computed. The process repeats until the value of $\boldsymbol{\Phi}_k$ converges. The complexity of the algorithm in an iteration is determined using the dominant matrix operations in Steps 1, 2, and 3 (cf. Fig. 2) which is: $O((Q+1)^2) + O(K^2 S)$.

III. POST-WEIGHTING SCHEMES AND OPTIMIZATION

To investigate the PW processing, first, we express z_k^φ in (8) as a function of s_k. By substituting (1b) into (8) and separating the first basis function from the rest, we get:

$$z_k^\varphi = \boldsymbol{h}_k^{\varphi T} \boldsymbol{W}_k \tilde{\boldsymbol{\phi}}_k^0 \phi_{k0}^0 s_k + \boldsymbol{h}_{k0}^\varphi{}^T \boldsymbol{W}_k \tilde{\boldsymbol{\phi}}_k^0 \boldsymbol{\Psi}'(s_k, c_k) \boldsymbol{\Phi}_k' + \tilde{Z}_{k,NL}^\varphi, \quad (11)$$

where $\boldsymbol{W}_k = \text{diag}(\boldsymbol{w}_k)$, $\tilde{\boldsymbol{\phi}}_{k0}^0 = [\phi_{k10}^0, \cdots, \phi_{kS0}^0]^T$, $\boldsymbol{\Psi}'(s_k, c_k) = \{\boldsymbol{\Psi}(s_k, c_k) \backslash \psi_0^0(s_k) = s_k\}$, $\boldsymbol{\Phi}_k' = \{\boldsymbol{\Phi}_k \backslash \phi_{k0}^0\}$, and $\tilde{Z}_{k,NL}$ is the nonlinear higher order terms with less power contents, obtained after removing the term consisting first (linear) basis function in (4). In (11), the first term is the desired output. But, the second and third terms inject nonlinearties, denoted as the nonlinear radiation $z_{k,NL}^\varphi$ as:

$$z_{k,NL}^\varphi = \boldsymbol{h}_k^{\varphi T} \boldsymbol{W}_k \tilde{\boldsymbol{\phi}}_k^0 \boldsymbol{\Psi}'(s_k, c_k) \boldsymbol{\Phi}_k' + \tilde{Z}_{k,NL}^\varphi. \quad (12)$$

In (12), $z_{k,NL}^\varphi$ gives the nonlinearties in other directions except φ. Because, in BO-DPD, its coefficients $\boldsymbol{\Phi}_k'$ are trained to provide the linearization in the direction φ. Therefore, to linearize it further in other directions too, the predistortion output of the trained DPD is passed through the PW block by changing the modes of the switches ($\mathcal{S}_0$ is open and $\{\mathcal{S}_i\}$ are closed). Next, we describe the proposed two PW schemes.

A. Post-Weighting

The outputs of the DPD are the $Q+1$ basis functions (in $\boldsymbol{\Psi}(s_k, c_k)$) multiplied by the respective coefficients (in $\boldsymbol{\Phi}_k$) and they are inputted to the PW block. The two types of PW schemes: (a) FF-PW and (b) LC-PW are shown in Fig. 3 for $S = 4$ and $Q = 3$. The PW coefficients are only multiplied by the Q nonlinear outputs of the DPD comprising the nonlinear basis functions (in $\boldsymbol{\Psi}'(s_k, c_k)$) except the linear basis function $\psi_{p_i}^{v_i} \mathcal{C}_k^{v_i} = s_k$ for $p_i = v_i = 0$ (cf. Section II-B). Thereafter, the outputs of the PW block are distributed to the inputs of the PAs to get the desired linear output.

1) Fully-Featured PW (FF-PW): For instance, the FF-PW scheme in Fig. 3(a) is for $S = 4$ and $Q = 3$. Here, $Q = 3$ nonlinear outputs of the DPD are multiplied by the $S = 4$ different sets of PW coefficients to generate the resultant predistorted signals for the respective S PAs. For the generation of the predistorted signal to the lth ($l \in \{1, 2, 3, 4\}$) PA, the corresponding set of coefficients is $\{\gamma_{klp_1}^{v_1}, \gamma_{klp_2}^{v_2}, \gamma_{klp_3}^{v_3}\}$. These PW coefficients are multiplied by the respective DPD outputs given in the set $\{\phi_{kp_1}^{v_1} \psi_{p_1}^{v_1} \mathcal{C}_k^{v_1}, \phi_{kp_2}^{v_2} \psi_{p_2}^{v_2} \mathcal{C}_k^{v_2}, \phi_{kp_3}^{v_3} \psi_{p_3}^{v_3} \mathcal{C}_k^{v_3}\}$. Thereafter, the sum of the three multiplications along with the linear DPD output, $\phi_{k0}^0 s_k$ gives the lth predistorted signal. Therefore, the total number of PW coefficients, $N_\gamma^F = S \times Q = 12$ which are completely different from each other without any repetition and it provides the full DOF to further linearize the individual PA. However, in LC-PW, the PW coefficients are assigned adaptively (some coefficients repeat) to the PAs to reduce their number as described later. In general, the Q PW coefficients associated with the predistorted signal to the lth PA are represented as: $\gamma_{klp_1}^{v_1}, \cdots, \gamma_{klp_Q}^{v_Q}$, where $\gamma_{klp_i}^{v_i}$ is the PW coefficient multiplied by the ith nonlinear DPD output. Further, for the convenience in analysis, the PW coefficients of the S PAs are arranged in a vector as: $\boldsymbol{\gamma}_k \triangleq [\gamma_{k1p_1}^{v_1}, \cdots, \gamma_{kSp_1}^{v_1}, \cdots, \gamma_{k1p_Q}^{v_Q}, \cdots, \gamma_{kSp_Q}^{v_Q}]^T$, where the

coefficients multiplied by the same DPD output are grouped together. The PW processing can be analyzed by incorporating the $\boldsymbol{\gamma}_k$ in (12). As its first term has $\boldsymbol{\Psi}'(s_k, c_k)$, to multiply $\boldsymbol{\gamma}_k$ with it, we need to rearrange $\boldsymbol{\Psi}'(s_k, c_k)$ along with other matrices and vectors according to arrangement of PW coefficients in $\boldsymbol{\gamma}_k$. So, $\boldsymbol{h}_k^\varphi$, $\boldsymbol{W}_k$, $\tilde{\boldsymbol{\phi}}_{k0}^0$, $\boldsymbol{\Psi}'(s_k, c_k)$, and $\boldsymbol{\Phi}_k'$ are arranged as $\overline{\boldsymbol{h}}_k^\varphi$, $\overline{\boldsymbol{W}}_k$, $\overline{\tilde{\boldsymbol{\phi}}}_{k0}^0$, $\overline{\boldsymbol{\Psi}}'(s_k, c_k)$, and $\overline{\boldsymbol{\Phi}}'_k$, respectively. $\overline{\boldsymbol{h}}_k^\varphi = \boldsymbol{1}_Q \otimes \boldsymbol{h}_k^\varphi$, $\overline{\boldsymbol{W}}_k = \text{diag}(\boldsymbol{1}_Q \otimes \boldsymbol{w}_k)$, $\overline{\tilde{\boldsymbol{\phi}}}_{k0}^0 = \text{diag}(\boldsymbol{1}_Q \otimes \tilde{\boldsymbol{\phi}}_{k0}^0)$, $\overline{\boldsymbol{\Psi}}'(s_k, c_k) = \text{diag}(\boldsymbol{\Psi}'(s_k, c_k)^T \otimes \boldsymbol{1}_S)$, and $\overline{\boldsymbol{\Phi}}'_k = \text{diag}(\boldsymbol{\Phi}_k' \otimes \boldsymbol{1}_S)$. Now, (12) can be expressed as:

$$\overline{z}_{k,NL}^\varphi = T_k^\varphi \boldsymbol{\gamma}_k + Z_{k,NL}^\varphi, \quad (13)$$

where $T_k^\varphi = \overline{\boldsymbol{h}}_k^{\varphi T} \overline{\boldsymbol{W}}_k \overline{\tilde{\boldsymbol{\phi}}}_{k0}^0 \overline{\boldsymbol{\Psi}}'(s_k, c_k) \overline{\boldsymbol{\Phi}}'_k$. Also, after incorporating PW coefficients in $\tilde{Z}_{k,NL}^\varphi$, it is expressed as $Z_{k,NL}^\varphi$ in (13). As described earlier, due to small power content, we neglect the PW effect on $Z_{k,NL}^\varphi$. Moreover, in FF-PW, the number of multipliers N_γ^F and the number of adders N_a^F are same as the number of PW coefficients in $\boldsymbol{\gamma}_k$, given as: $N_\gamma^F = N_a^F = S \times Q$. Also, the number of RF chains $N_{RF}^F = S$. So, to reduce them, we propose a LC-PW scheme.

2) Low-Complexity PW (LC-PW): As the complexity of the PW block depends on the number of PW coefficients, therefore, in the LC-PW scheme, the number of PW coefficients are reduced by a factor based on a geometric sequence. In general, the nonlinear DPD outputs in this scheme are arranged in their decreasing order of dominance (or increasing order of the polynomial terms). Besides, the numbers of PW coefficients that are multiplied by the Q DPD outputs, decrease in a geometric sequence as: $\{Sr^\nu, Sr^{(\nu+1)}, \cdots, Sr^{(\nu+Q-1)}\}$, where r (< 1) is the common ratio. For example, in Fig. 3(b), for $S = 4$, $Q = 3$, $r = 1/2$, and $\nu = 1$, the sequence of numbers of PW coefficients is: $\{2, 1, 0.5\}$. But, the number cannot be a fraction value, so, the value in the sequence less than one is assigned as one, thus the sequence is: $\{2, 1, 1\}$. So, we keep decreasing the number of coefficients until $Sr^i < 1$; $i \in \{\nu, \nu+1, \cdots, \nu+Q-1\}$ and after that we assign one PW coefficient to each of the remaining DPD outputs. Thus, in general, the total number of coefficients (or the multipliers), N_γ^L in the LC-PW is given by (14a) and the total number of adders, N_a^L is expressed in (14b).

$$N_\gamma^L = \begin{cases} S \times \dfrac{r^\nu(1-r^Q)}{1-r}; & \text{for } S \times r^{(Q+\nu-1)} \geq 1 \\ S \times \dfrac{r^\nu(1-r^m)}{1-r} + Q - m; & \text{for } \{S \times r^{(m+\nu-1)} \geq 1\} \\ & \wedge \{S \times r^{(m+\nu)} < 1\}, \end{cases} \quad (14a)$$

$$N_a^L = N_\gamma^L + Sr^\nu - \lceil Sr^{(\nu+Q-1)} \rceil. \quad (14b)$$

For the first case in (14a), the PW coefficient vector, $\boldsymbol{\gamma}_k = [\gamma_{k1p_1}^{v_1}, \gamma_{k(1+r^{-\nu})p_1}^{v_1}, \cdots, \gamma_{k(1+(Sr^\nu-1)r^{-\nu})p_1}^{v_1}, \cdots, \gamma_{k1p_Q}^{v_Q}, \gamma_{k(1+r^{-(\nu+Q-1)})p_Q}^{v_Q}, \cdots, \gamma_{k(1+(Sr^{(\nu+Q-1)}-1)r^{-(\nu+Q-1)})p_Q}^{v_Q}]^T$ and $\overline{\boldsymbol{\Phi}}'_k = \text{diag}(\boldsymbol{\Phi}_k' \otimes \boldsymbol{1}_S)\text{diag}(\underbrace{\boldsymbol{1}_{r^{-\nu}}, \cdots, \boldsymbol{1}_{r^{-\nu}}}_{Sr^\nu \text{ times}}, \cdots, \underbrace{\boldsymbol{1}_{r^{-(\nu+Q-1)}}, \cdots, \boldsymbol{1}_{r^{-(\nu+Q-1)}}}_{Sr^{(\nu+Q-1)} \text{ times}})$. Otherwise, for the second case, $\boldsymbol{\gamma}_k = [\gamma_{k1p_1}^{v_1}, \gamma_{k(1+r^{-\nu})p_1}^{v_1}, \cdots, \gamma_{k(1+(Sr^\nu-1)r^{-\nu})p_1}^{v_1}, \cdots, \gamma_{k1p_m}^{v_m}, \gamma_{k(1+r^{-(\nu+m-1)})p_m}^{v_m}, \cdots, \gamma_{k(1+(Sr^{(\nu+m-1)}-1)r^{-(\nu+m-1)})p_m}^{v_m},$

$\gamma_{k1p_{(m+1)}}^{v(m+1)}, \gamma_{k1p_{(m+2)}}^{v(m+2)}, \cdots, \gamma_{k1p_Q}^{v_Q}]^T$ and $\overline{\Phi}'_k = \text{diag}(\Phi'_k \otimes \mathbf{1}_S)\text{diag}(\underbrace{\mathbf{1}_{r^{-\nu}}, \cdots, \mathbf{1}_{r^{-\nu}}}_{Sr^\nu \text{ times}}, \cdots, \underbrace{\mathbf{1}_{r^{-(\nu+m-1)}}, \cdots, \mathbf{1}_{r^{-(\nu+m-1)}}}_{Sr^{(\nu+m-1)} \text{ times}},$ $\underbrace{\mathbf{1}_S, \cdots, \mathbf{1}_S}_{Q-m \text{ times}})$. Moreover, remaining vectors and matrices, $\overline{h}_k^\varphi$,

$\overline{W}_k$, and $\overline{\tilde{\phi}}_{k0}^0$, $\overline{\Psi}'(s_k, c_k)$ are expressed same as for FF-PW (cf. Section III-A1). Also, the expression for the nonlinear radiation, $\overline{z}_{k,NL}^\varphi$ is same as in (13). Moreover, from Fig. 3(b), the number of RF chains, N_{RF}^L depends on the number of coefficients assigned to the first nonlinear output, i.e., $N_{RF}^L = Sr^\nu = N_{RF}^F r^\nu$. Thus, in LC-PW, the number of RF chains, multipliers, and the adders are reduced by the factors, $r^{-\nu}$, N_γ^F/N_γ^L, and N_a^F/N_a^L, respectively. For example, in Fig. 3(b), the respective factors are 2, 4, and 2.4. Hence, the LC-PW is less complex and economical for the mMIMO.

B. Optimization of γ_k

The optimization problem can be formulated as:

$\mathcal{P}_0: \underset{\gamma_k}{\text{minimize}} \quad \sum_t \mathbb{E}[|\overline{z}_{k,NL}^{\varphi_t}|^2]$

s. t.: $\overline{h}_k^{\varphi_0 T} \overline{W}_k \overline{\tilde{\phi}}_{k0}^0 \overline{\Psi}' \overline{\Phi}'_k \gamma_k = \overline{h}_k^{\varphi_0 T} \overline{W}_k \overline{\tilde{\phi}}_{k0}^0 \overline{\Psi}' \overline{\Phi}''_k,$

where $\overline{\Phi}''_k = \Phi'_k \otimes \mathbf{1}_S$ and $\mathbb{E}[\cdot]$ is the expectation with respect to time samples. In problem $\mathcal{P}_0$, the objective function which needs to be minimized in γ_k, is the sum of the average value of the power of nonlinear radiation in the given range of directions with sample points $\{\varphi_t\}$. The constraint ensures the linearization of BO signal in the desired direction φ_0.

1) Optimal Solution: To investigate the convexity of the problem, the objective function, $\mathcal{O}_{\mathcal{P}_0} \triangleq \sum_t \mathbb{E}[|\overline{z}_{k,NL}^{\varphi_t}|^2] = \sum_t \mathbb{E}[\overline{z}_{k,NL}^{\varphi_t H} \overline{z}_{k,NL}^{\varphi_t}]$, can be expressed using (13) as:

$$\mathcal{O}_{\mathcal{P}_0} = \gamma_k^H \sum_t \mathbb{E}[T_k^{\varphi_t H} T_k^{\varphi_t}]\gamma_k + \gamma_k^H \sum_t \mathbb{E}[T_k^{\varphi_t H} Z_{k,NL}^{\varphi_t}]$$
$$+ \sum_t \mathbb{E}[Z_{k,NL}^{\varphi_t H} T_k^{\varphi_t}]\gamma_k + \sum_t \mathbb{E}[Z_{k,NL}^{\varphi_t H} Z_{k,NL}^{\varphi_t}]. \quad (15)$$

As the constraint is linear and from (15), the objective function is quadratic in γ_k, the problem $\mathcal{P}_0$ is convex and gives a global solution using the Karush–Kuhn–Tucker (KKT) conditions. From $\mathcal{P}_0$, the Lagrangian function, $\mathcal{L}(\gamma_k, \eta)$ is:

$$\mathcal{L}(\gamma_k, \eta) = \mathcal{O}_{\mathcal{P}_0} + \eta(T_k^{\varphi_0}\gamma_k - \overline{h}_k^{\varphi_0 T} \overline{W}_k \overline{\tilde{\phi}}_{k0}^0 \overline{\Psi}' \overline{\Phi}''_k), \quad (16)$$

where η is the Lagrangian multiplier which is $\neq 0$ to consider the constraint in the optimization. Using the complex gradient of $\mathcal{L}(\gamma_k, \eta)$ in γ_k, the KKT conditions are obtained as:

$$\sum_t \mathbb{E}[T_k^{\varphi_t H} T_k^{\varphi_t}]\gamma_k + \sum_t \mathbb{E}[T_k^{\varphi_t H} Z_{k,NL}^{\varphi_t}] + \eta T_k^{\varphi_0 H} = 0, \quad (17a)$$

$$T_k^{\varphi_0}\gamma_k = T_k^{\varphi'_0}, \quad (17b)$$

where $T_k^{\varphi'_0} = \overline{h}_k^{\varphi_0 T} \overline{W}_k \overline{\tilde{\phi}}_{k0}^0 \overline{\Psi}' \overline{\Phi}''_k$. Using (17), the optimal solution $\widehat{\gamma}_k$ and corresponding Lagrangian multiplier $\widehat{\eta}$ are:

$$\widehat{\gamma}_k = \widehat{\eta}\left(\sum_t \mathbb{E}[T_k^{\varphi_t H} T_k^{\varphi_t}]\right)^{-1} T_k^{\varphi_0 H} - \left(\sum_t \mathbb{E}[T_k^{\varphi_t H} T_k^{\varphi_t}]\right)^{-1}$$
$$\times \sum_t \mathbb{E}[T_k^{\varphi_t H} Z_{k,NL}^{\varphi_t}], \quad (18a)$$

$$\widehat{\eta} = \frac{T_k^{\varphi'_0} + T_k^{\varphi_0}(\sum_t \mathbb{E}[T_k^{\varphi_t H} T_k^{\varphi_t}])^{-1} \sum_t \mathbb{E}[T_k^{\varphi_t H} Z_{k,NL}^{\varphi_t}]}{T_k^{\varphi_0}(\sum_t \mathbb{E}[T_k^{\varphi_t H} T_k^{\varphi_t}])^{-1} T_k^{\varphi_0 H}}. \quad (18b)$$

Fig. 4. Performance comparison of the different PW schemes against (a) the benchmark schemes and (b) the systems with only BO-DPD.

TABLE I
AVERAGE ACPR OF TWO ADJACENT CHANNELS.

Scheme	ACPR (dB)	Scheme	ACPR (dB)
Intra [6]	41.6683	BO-DPD w/ CTP [2]	33.0106
Inter [6]	45.7127	BO-DPD w/o CTP [4]	27.5087
FF-PW	63.4938	LC-PW	52.1726

IV. NUMERICAL EXPERIMENT AND CONCLUSION

To evaluate the performance, we use the set of 16 PA memoryless polynomial models obtained by measuring the outputs of 16 HMC943APM5E PA ICs at 28.5 GHz to the OFDM input signal of 200 MHz bandwidth. Although, the signal is wideband, the memoryless model provides an accuracy around -23 dB in normalized mean square error [3]. Besides, two subarrays, each comprising 16 PAs, are considered. The PAs are arranged in a uniform linear array and the distance between the two adjacent antennas is half the operating wavelength. The beamforming and steering weight for a given azimuth angle is determined using the procedure in [7]. The crosstalk between two adjacent antennas is -10 dB, whereas, it decays as a square of distance between other two antennas [3]. Using the measured output of HMC943APM5E PA ICs at 28.5 GHz with -10 dB crosstalk, we identified the dual-input memoryless model for each PA using LS estimation as follows. For a given input signal $\mathring{x}_k$, a test crosstalk signal $\mathring{c}_{kl}$ to the PA, and the measured output $\mathring{y}_{kl}$, the coefficients of the PA are identified using LS estimation as: $\widehat{\Phi}_{kl} = \Psi(\mathring{x}_k, \mathring{c}_{kl})^\dagger \mathring{y}_{kl}$.

Fig. 4 depicts the performance comparison of the schemes in the nonlinear radiations. A notch in each curve at 0 rad is due to BO-DPD linearization in the direction. PW coefficients are optimized for the angle range $[-\pi/3, \pi/3]$ and $[-\pi/2, \pi/2]$ for Figs. 4(a) and 4(b), respectively. So, intra-PW and inter-PW schemes overshoots beyond the range in Fig. 4(a). In the direct nonlinear radiation (DNR), the message signal is directly transmitted from the subarray where the average power of the transmit signal is normalized to 0 dB. So, the predistortion schemes give significant performance enhancement against DNR. Further, FF-PW and LC-PW provide the improvements by 20.01 dB and 9.88 dB against intra-PW due to their higher DOF in PW, but, inter-PW gives marginal improvement of 3.44 dB. Fig 4(b) quantifies the performance improvement due to crosstalk preprocessing (CTP) in the schemes. Here, a scheme name followed by w/ (w/o) represents the predistortion with (without) CTP. In BO-DPD w/o [4] and BO-DPD w/ [2],

only DPD is used to linearize the BO outputs. Thus, the PW schemes always perform better. Moreover, BO-DPD, LC-PW, and FF-PW schemes with CTP provide the respective on average improvements by 5.06 dB, 5.82 dB, and 8.11 dB against the schemes without CTP. Also, the average adjacent channel power ratios (ACPRs) of the schemes are shown in Table I. Here, average ACPR is the average of the ACPRs of the two adjacent channels which are computed as in [3]. Again, the FF-PW has the best average ACPR while the LC-PW has an intermediate performance.

REFERENCES

[1] H. Yan and D. Cabric, "Digital predistortion for hybrid precoding architecture in millimeter-wave massive MIMO systems," in *Proc. IEEE ICASSP*, LA, USA, Mar. 2017, pp. 3479–3483.

[2] Q. Luo, X.-W. Zhu, C. Yu, and W. Hong, "Single-receiver over-the-air digital predistortion for massive MIMO transmitters with antenna crosstalk," *IEEE Trans. Microw. Theory Tech.*, vol. 68, no. 1, pp. 301 – 315, Jan. 2019.

[3] A. Brihuega *et al.*, "Digital predistortion for multiuser hybrid MIMO at mmwaves," *IEEE Trans. Signal Process.*, vol. 68, pp. 3603 – 3618, May 2020.

[4] E. Ng, Y. Beltagy, P. Mitran, and S. Boumaiza, "Single-input single-output digital predistortion of power amplifier arrays in millimeter wave RF beamforming transmitters," in *Proc. IEEE Int. Microw. Symp.-IMS*, Philadelphia, PA, USA, Jun. 2018, pp. 481–484.

[5] C. Yu *et al.*, "Full-angle digital predistortion of 5G millimeter-wave massive MIMO transmitters," *IEEE Trans. Microw. Theory Tech.*, vol. 67, no. 7, pp. 2847–2860, Jul. 2019.

[6] J. Yan, H. Wang, and J. Shen, "Novel post-weighting digital predistortion structures for hybrid beamforming systems," *IEEE Commun. Lett.*, vol. 25, no. 12, pp. 3980–3984, Dec. 2021.

[7] B. Khan *et al.*, "Statistical digital predistortion of 5G millimeter-wave RF beamforming transmitter under random amplitude variations," *IEEE Trans. Microw. Theory Tech.*, vol. 70, no. 9, pp. 4284–4296, Sep. 2022.