Digital Predistortion for Multiuser Hybrid MIMO at mmWaves

Alberto Brihuega, Student Member, IEEE, Lauri Anttila, Member, IEEE, Mahmoud Abdelaziz, Member, IEEE, Fredrik Tufvesson, Fellow, IEEE, and Mikko Valkama, Senior Member, IEEE

Abstract—Efficient mitigation of power amplifier (PA) nonlinear distortion in hybrid precoding based broadband mmWave systems is an open research problem. In this article, we first carry out detailed signal and distortion modeling in broadband multi-user hybrid MIMO systems with a bank of nonlinear PAs in each subarray. Building on the derived models, we then propose a novel digital predistortion (DPD) solution that requires only a single DPD unit per transmit chain or subarray. The proposed DPD system makes use of a closed-loop learning architecture and combined feedback observation receivers that merge the individual PA output signals within each subarray for DPD parameter learning purposes. Such combined feedback signals reflect the true received signals at the intended users, from the nonlinear distortion point of view. We show that, under spatially correlated multipath propagation, each DPD unit can provide linearization toward every intended user, or more generally, towards all spatial directions where coherent propagation is taking place. In the directions with less coherent combining, the joint effect of DPD and beamforming keeps the nonlinear distortion at a sufficiently low level. Extensive numerical results are provided, demonstrating and verifying the excellent linearization performance of the proposed DPD system in different evaluation scenarios.

Index Terms—Digital predistortion, millimeter wave communications, large-array transmitters, hybrid MIMO, multi-user MIMO, frequency-selective channels, power amplifiers, nonlinear distortion, out-of-band emissions.

I. INTRODUCTION

THE demands for higher data rates and larger network capacities have led mobile communications system evolution to adopt new spectrum at different frequency bands, to deploy larger and larger antenna arrays, and to substantially densify the networks [1]–[6]. In the lower frequency bands, specifically the so-called sub-6 GHz region, very aggressive spatial multiplexing [7], [8] is one key technology. In such systems, it is common to assume that spatial precoding or beamforming can be done primarily digitally, offering the maximum flexibility to select and optimize the precoder weights, compared to analog beamforming that is subject to maximum flexibility to select and optimize the precoder weights, compared to analog beamforming that is subject to multiple physical constraints [6], [9], [10]. Millimeter wave (mmWave) communications, on the other hand, allow to leverage the large amounts of available spectrum in order to provide orders of magnitude higher data rates, but also impose multiple challenges compared to sub-6 GHz systems. In general, the propagation losses at mmWaves are considerably higher than those at sub-6 GHz bands, and thus large antenna gains are typically needed at both the transmitter and receiver ends in order to facilitate reasonable link budgets [1]–[4], [11].

Operating at mmWave frequencies allows to pack a large number of antennas in a small area. However, the implementation of fully digital beamforming based large antenna array transmitters turns out to be very costly and power consuming [12]. For this reason, many works have proposed and considered hybrid analog-digital beamforming solutions [9]–[17] as a feasible technical approach and compromise between implementation costs, power consumption, and beamforming flexibility. This is also well in-line with the angular domain sparsity of the mmWave propagation channels [4], [10], [17], [18], which results in reduced multiplexing gain. In general, there are several hybrid architectures depending on how the analog beamforming stage is implemented [11], [12]. Two common architectures are the so-called full-complexity architecture, where an individual analog precoder output is a linear combination of all the RF signals, and the so-called reduced-complexity architecture, in which each TX chain is connected only to a subset of antennas, known as subarray. The reduced complexity architecture, illustrated in Fig. 1, is known to be more feasible for real implementations [11], [12], [14]–[16], [19] and is thus assumed also in this article.

A. Nonlinear Distortion and State-of-the-Art

In general, energy efficiency is an important design criterion for any modern radio system, including 5G and beyond cellular systems [1], [2]. Therefore, in the large array transmitter context, efficient operation of the power amplifier (PA) units is of key importance. To this end, highly nonlinear PAs operating close to saturation are expected to be used in the base stations (BS) [20]. Nonlinear distortion due to PAs in massive MIMO transmitters has been studied in the recent literature [21]–[28]. In [27], the out-of-band (OOB) emissions due to nonlinear PAs were analyzed in single antenna and multiantenna transmitters, considering both line-of-sight (LOS) and non-line-of-sight (NLOS) propagation, and assuming different memoryless polynomial models per antenna branch. It was shown that the adjacent channel leakage ratio (ACLR) in multiantenna transmitters when serving a single user is, in the worst case, at the same level as in single-antenna transmitters when both systems provide the same received signal power. The worst
case emissions occur in the direction of the intended receiver, regardless of LOS or NLOS propagation, since OOB emissions also get beamformed towards this direction, while in other directions they get diluted due to less coherent superposition. Understanding the spatial characteristics of the unwanted emissions is of fundamental importance, since the neighboring channel emissions can even violate the spurious emission limits as demonstrated in [23].

Compared to simply backing off the PA input power, a much more efficient approach to control the PA-induced emissions while still operating close to saturation is to utilize digital predistortion (DPD) [29], [30]. DPD has been recently studied in the context of large antenna arrays in [31]–[40]. In [31], [32], fully digital beamforming based system was investigated. In [31], a dedicated DPD unit per antenna/PA was considered, primarily focusing on the reduction of the complexity of the DPD learning algorithm. However, a dedicated DPD unit per antenna/PA branch may not be implementation-feasible in large array transmitters because of the complexity and power consumption issues. Therefore, in [32] the authors proposed an alternative DPD solution where a single DPD unit can linearize an arbitrarily large antenna array, with multiple PAs, when single-user phase-only digital precoding is considered.

In [33], [35]–[39], DPD solutions for single-user hybrid MIMO transmitters were investigated assuming the reduced-complexity architecture shown in Fig. 1. To this end, and since each DPD unit operates in the digital domain, an individual predistorter is responsible for linearizing all the PAs within its respective subarray. Since the PA units are in practice mutually different, this is essentially an under-determined problem and generally yields reduced linearization performance, when compared to linearizing each PA individually. In [33], the DPD learning is based on observing only a single PA output, within each subarray, while the works in [34], [40] consider the multiuser case but adopt a simplifying assumption that all the PAs are mutually identical. As a result, both approaches lead to reduced linearization performance in practice, due to the mutual differences between real PA units and their exact nonlinear distortion characteristics. Additionally, only a third-order PA model and corresponding DPD processing are considered in [34].

The most recent works [36]–[39] seek to benefit from the spatial characteristics of the OOB emissions in array transmitters in order to develop efficient DPD solutions. These works rely on the fact that unwanted emissions are more significant in the direction of the intended receiver, while emissions in other spatial directions are attenuated by the array response, as explained in [27]. In the single-user case, the received signal of the intended user under LOS propagation can be mimicked by coherently combining all the individual PA output signals within the subarray. This forms the signal for DPD parameter learning and overall effectively yields a well defined single-input-single-output DPD problem. Such DPD processing results in minimizing the OOB emissions in the direction of the intended receiver [36], The works in [32]–[40] either assume single-user transmission or adopt some other simplifying assumptions such as all PAs being identical, pure LOS propagation or narrowband fading. Thus, DPD techniques for true multi-user hybrid MIMO systems under mutually different PA units and broadband channels do not exist in the current literature.

B. Novelty and Contributions

In this paper, we first provide detailed signal and distortion modeling for hybrid-precoded multi-user MIMO systems under nonlinear PAs. Building on the derived models, we then propose a novel DPD solution for efficient mitigation of PA nonlinearities such that only a single DPD unit per TX chain or subarray is deployed. In general, due to hybrid precoding and multi-user transmission, the received signals by the intended and potential victim users are contributed by the transmission from all the subarrays. As a consequence, the overall DPD system needs to provide linearization not only to a single point in space, as was the case in [36], [37], but to multiple points and corresponding receivers. To this end, considering that unwanted emissions in array transmitters are strongest in the directions of the intended receivers, we primarily focus on reducing the inband and out-of-band emissions in these directions, while rely on the joint effects of beamforming and DPD processing in other directions. For parameter estimation purposes, the PA output signals, per each subarray, are coherently combined in the RF domain in order to generate the feedback signals for the closed-loop adaptive learning system, requiring only a single observation receiver per TX chain. The resulting combined signals reflect the actual nonlinear distortion radiated from each subarray, while the composite nonlinear distortion observed by the intended receivers is suppressed by the overall DPD system. Specifically, we show that under spatially correlated multipath propagation, within a subarray, each DPD unit can provide linearization towards every intended user, or more generally, towards all spatial directions where coherent propagation is taking place. For the directions with less coherent combining, it is shown that the joint effect of DPD and beamforming keeps the nonlinear distortions at a low level.
The remainder of this paper is organized as follows: In Section II, the hybrid multiuser MIMO system model considered in this work is described. In Section III, the modeling and analysis of the nonlinear distortion arising from the nonlinear PAs are carried out, with specific emphasis on the combined or observable distortion. Then, Section IV describes the proposed DPD structure and parameter learning solution. In Section V, the numerical performance evaluation results are presented and comprehensively analyzed. Lastly, Section VI will provide the main concluding remarks.

II. MULTIUSER HYBRID MIMO SYSTEM MODEL

A. Basics

The overall considered hybrid beamforming based multiuser MIMO-OFDM transmitter is shown in Fig. 2, containing \( L \) TX chains and \( M \) antenna units per subarray, while serving \( U \) single-antenna users simultaneously. The subcarrier-wise BB precoder is responsible for mapping the \( U \) data streams onto \( L \) TX chains and for spatially multiplexing the different users, while the RF precoder focuses the energy towards the dominant directions of the channel. It is further assumed that \( U \leq L \leq LM \).

The samples of the \( U \) data streams at the \( k \)-th subcarrier, expressed as \( s[k] = (s_1[k], s_2[k], \ldots, s_U[k])^T \), are first digitally precoded by means of the precoder matrix \( F[k] \in \mathbb{C}^{L \times U} \) yielding the precoded data vector \( x[k] = F[k]s[k] \in \mathbb{C}^{L \times 1} \). The design and optimization of the BB precoder weights in hybrid beamforming system can, in general, be done in multiple different ways [9], [10], [17], while our assumptions are shortly described in Subsection II-C. The precoded data symbol blocks are then transformed to time-domain waveforms through IFFT's of size \( K_{\text{FFT}} > K_{\text{ACT}} \) where \( K_{\text{ACT}} \) denotes the number of active subcarriers. A cyclic prefix of length \( K_{\text{CP}} \) is then added to the sample blocks. The basic system model also contains peak-to-average-power ratio (PAPR) reduction to improve the power efficiency of the transmitter, as well as windowing to obtain better spectral containment for the OFDM signals. After these operations, the \( L \) signals are mapped onto their respective antenna branches by means of the analog precoder, expressed as a matrix \( W \in \mathbb{C}^{M_{\text{TOT}} \times L} \), where \( M_{\text{TOT}} = LM \) stands for the total number of antenna units in the transmitter. Overall, when interpreted at subcarrier \( k \), this yields a precoded vector of the form

\[
v[k] = WF[k]s[k]. \tag{1}
\]

As the analog precoder operates in time-domain, typically in the form of simple phase-rotators, it is common to all the subcarriers.

In the over-the-air propagation, again interpreted at subcarrier \( k \), the samples \( v[k] \in \mathbb{C}^{M_{\text{TOT}} \times 1} \) effectively combine through the frequency-selective array channels towards the receiving devices. Denoting the array channel of the \( u \)-th user at subcarrier \( k \) by \( g_u[k] \in \mathbb{C}^{M_{\text{TOT}} \times 1} \), and assuming that the cyclic prefix is longer than the channel delay spread, the corresponding received signal model reads

\[
z_u[k] = g_u^T[k]WF[k]s[k] + n_u[k], \tag{2}
\]

where \( n_u[k] \sim \mathcal{N}(0, \sigma_n^2) \) refers to additive Gaussian noise.

B. mmWave Channel Model

In order to accurately incorporate the frequency-selectivity as well as the spatial correlation characteristics of the array channels, we adopt a geometry-based clustered modeling approach, similar to [9], [17], [19]. Specifically, we assume a clustered channel model with \( C \) clusters, where each cluster is made up of \( R \) rays. Each cluster \( c \) has a certain path-delay \( \tau_c \) and angle of arrival \( \theta_c \), while each ray has its corresponding ray-delay and angle of arrival denoted by \( \tau_r \) and \( \phi_r \), respectively. The corresponding angles of departure of the paths and rays from each cluster to each user are denoted by \( \gamma_c \) and \( \phi_r \), respectively. Lastly, let \( f_c(n) \) denote a \( T_s \) spaced raised-cosine pulse shaping function evaluated at the time instant \( n \). Following the above mentioned model, the delay-d channel vector [9] for the \( u \)-th user reads then

\[
h_u[d] = \sum_{c=1}^{C} \sum_{r=1}^{R} h_r f_c (dT_s - \tau_c - \tau_r) a_{\text{Rx}} (\gamma_c - \phi_r) a_{\text{Tx}} (\theta_c - \phi_r), \tag{3}
\]

where \( h_r \) is the complex gain corresponding to the \( r \)-th ray and is drawn from a zero-mean-unit-variance circular symmetric Gaussian distribution, \( a_{\text{Rx}} \) denotes the response of the TX array [15], [16], [41], while \( a_{\text{Rx}} \) accounts for the phase between the clusters and the user. The corresponding delay-d multiuser MIMO channel matrix reads then

\[
H[d] = (h_1[d], h_2[d], \ldots, h_U[d])^T \in \mathbb{C}^{U \times M_{\text{TOT}}}. \tag{4}
\]

Finally, the corresponding multiuser frequency-domain response at subcarrier \( k \), denoted by \( G[k] = (g_1[k], g_2[k], \ldots, g_U[k])^T \in \mathbb{C}^{U \times M_{\text{TOT}}} \), is given by

\[
G[k] = \sum_{d=0}^{D-1} H[d] e^{-j \frac{2 \pi d k}{M_{\text{TOT}}}}. \tag{4}
\]

A LOS component can also be added, on top of the channel model in (3), in order to account for Ricean fading with any given Ricean K-factor defined as the power ratio between the received LOS and NLOS components [42].

C. Design of Digital and Analog Precoders

The design and optimization of the digital and analog precoders in hybrid MIMO transmitters is generally a challenging problem [11], [12] for several reasons. The analog and digital precoders constitute a cascaded system, therefore, both blocks are coupled making the resulting optimization problem non-convex [9], [10], [12], [17]. Furthermore, since the analog precoders are typically implemented as a network of phase shifters, this imposes additional constraints, such as having a limited set of available phase rotations. One common approach is thus to decouple the design of the baseband and analog precoders. The analog precoder can be first selected based on beamsteering the signals towards the dominant directions of the channel, while the BB precoding, that acts over the equivalent channel (analog precoder and actual channel response), is responsible for reducing the multi-user interference and compensating for the frequency-selectivity of the channel.

Provided that the analog precoder is known or fixed, the BB precoding matrix at the \( k \)-th subcarrier can be obtained in a straight-forward manner, by utilizing the equivalent channel
matrix $G_{eq}[k] = G[k]W$. For example, the zero-forcing (ZF) and regularized ZF (RZF) precoders essentially read [8, 43]

$$F_{ZF}[k] = G_{eq}^H[k] (G_{eq}[k] G_{eq}^H[k])^{-1}$$
$$F_{RZF}[k] = G_{eq}^H[k] (G_{eq}[k] G_{eq}^H[k] + \delta I)^{-1}. \quad (5)$$

For transmit power normalization, additional scaling factors can be introduced, building on, e.g., a sum-power constraint [9, 17, 19].

For the reduced-complexity architecture, the composite analog precoder matrix is in general of the form

$$W = \begin{bmatrix}
    w_1 & 0 & \cdots & 0 \\
    0 & w_2 & \cdots & 0 \\
    \vdots & \ddots & \vdots & \vdots \\
    0 & 0 & \cdots & w_L
\end{bmatrix}, \quad (7)$$

where $w_l = (w_{l,1}, w_{l,2}, \ldots, w_{l,M})^T \in \mathbb{C}^{M \times 1}$ is the beamforming vector of the $l$-th subarray. Assuming further that the analog precoder coefficients $w_{l,m}$ are simply phase-rotations, $|w_{l,m}| = 1 \ \forall l,m$. Interestingly, the phase rotators $w_{l,m}$ can be optimized in multiple ways, while we conceptually differentiate between the following two main alternatives:

1) **Single-beam analog beamformer**: A subarray generates a single beam towards the main channel tap of a particular user. An individual user is then being primarily served by a single subarray. It is, however, important to note that the actual received signal of every user is still contributed by the transmitted signals of all the subarrays since practical beampatterns provide only limited spatial isolation.

2) **Multi-beam analog beamformer**: Each subarray generates multiple beams, one per user, simultaneously. All the users are then more evenly served by all the subarrays, and thus the received signals are not dominated by the transmissions from a single subarray. In order to generate multiple simultaneous beams through phase-only precoding, one can refer, e.g., to [44]. In general, the multi-beam approach per subarray is more natively reflecting true multiuser hybrid beamforming.

III. MODELING AND ANALYSIS OF PA NONLINEAR DISTORTION

To build the basis for the actual DPD developments, the modeling of the PA-induced nonlinear distortion is next pursued, with specific emphasis on the observable or combined distortion at receiver end. Similar to [32], [35], and for presentation convenience, we consider memoryless polynomial based PA models in the analysis. Additionally, different PA units are mutually different, no DPD processing is yet considered, and all modeling is carried out in discrete-time baseband equivalent domain.

Now, consider the $m$-th antenna branch in the $l$-th subarray, and let $v_{l,m}(n) = w_{l,m} x_l(n)$ denote the PA input signal where $w_{l,m}$ refers to the analog beamformer weight while $x_l(n)$ denotes the digitally precoded sample sequence of the $l$-th TX.

The corresponding PA output signal can then be expressed as

$$y_{l,m}(n) = \sum_{p=1}^{P} \alpha_{l,m,p} v_{l,m}(n) |v_{l,m}(n)|^{p-1}$$
$$= \sum_{p=1}^{P} w_{l,m} \alpha_{l,m,p} x_l(n) |w_{l,m} x_l(n)|^{p-1}, \quad (8)$$

where $\alpha_{l,m,p}$ stands for the $p$-th order PA coefficient at the $m$-th antenna branch of the subarray $l$ while $P$ is the corresponding polynomial order. Since $|w_{l,m}| = 1$, the PA output signal can be re-written as

$$y_{l,m}(n) = w_{l,m} \sum_{p=1}^{P} \alpha_{l,m,p} x_l(n) |x_l(n)|^{p-1} \quad (9)$$
$$= w_{l,m} \sum_{p=1}^{P} \alpha_{l,m,p} \psi_{l,p}(n), \quad (10)$$
where \( \psi_{l,p}(n) = x_l(n)|x_l(n)|^{p-1} \) denotes the so-called static nonlinear (SNL) basis function of order \( p \).

Let us next consider the observable combined signal at user \( u \), being contributed by all antenna elements of all subarrays. Denoting the impulse response between the \( m \)-th antenna element of the \( l \)-th subarray and the \( u \)-th user by \( h_{l,m,u}(n) \), the received signal excluding additive thermal noise for notational simplicity reads

\[
z_u(n) = \sum_{l=1}^{L} \sum_{m=1}^{M} h_{l,m,u}(n) \ast \sum_{p=1}^{P} w_{l,m} \alpha_{l,m,p} \psi_{l,p}(n), \tag{11}
\]

where \( \ast \) is the discrete-time convolution operator. It can be observed from (11) that the composite received signal is of a Hammerstein \([45]–[47]\) form, with different tap delays introduced by the multipath channels. Assuming next that the individual channels within a single subarray are clearly correlated, a common assumption at mmWaves \([17], [19]\), one can argue that \( h_{l,m,u}(n) \approx h_{l,u}(n)e^{j\beta_{l,m,u}} \), and thus rewrite (11) as

\[
z_u(n) = \sum_{l=1}^{L} h_{l,u}(n) \ast \sum_{m=1}^{M} \sum_{p=1}^{P} e^{j\beta_{l,m,u}} w_{l,m} \alpha_{l,m,p} \psi_{l,p}(n), \tag{12}
\]

where \( e^{j\beta_{l,m,u}} \) stems from the phase differences between the signals due to the array geometry as well as exact propagation conditions. Furthermore, for notational convenience, the phase of the dominant channel tap of \( h_{l,u}(n) \) is assumed to be embedded in \( e^{j\beta_{l,m,u}} \). Such an approximation is well-argued at mmWaves, where it is typically a dominating LOS path and only few scatterers \([17], [19]\). The assumption naturally holds also under pure LOS scenario, as well as under geometric channel models with small antenna spacing such that the spatial correlation is high. It is important to note, however, that the channels between subarrays are considered to be already substantially less correlated, in general.

In order to have a better insight into the structure of the observable nonlinear distortion, we focus next on the received signals of two users, say \( u \) and \( u' \), and specifically investigate the contribution of the \( l \)-th TX chain only, expressed as

\[
z_{u}^{l}(n) = h_{l,u}(n) \ast \sum_{m=1}^{M} \sum_{p=1}^{P} e^{j\beta_{l,m,u}} w_{l,m} \alpha_{l,m,p} \psi_{l,p}(n) \tag{13}
\]

\[
z_{u'}^{l}(n) = h_{l,u'}(n) \ast \sum_{m=1}^{M} \sum_{p=1}^{P} e^{j\beta_{l,m,u'}} w_{l,m} \alpha_{l,m,p} \psi_{l,p}(n) \tag{14}
\]

Now, it can be seen from (13) and (14) that the received signals at different receivers, stemming from a given subarray, have a very similar structure. The nonlinear terms are shaped by the same analog precoder coefficients and the same PA responses, while only the channel impulse responses and the element-wise phase differences differ. Then, by considering the multi-beam analog beamformer discussed in Section II-C, for generality purposes and to harness true multi-user hybrid MIMO, coherent combining towards both users can be achieved, and hence, (13) and (14) can be re-written as

\[
z_{u}^{l}(n) = h_{l,u}(n) \ast \sum_{m=1}^{M} \sum_{p=1}^{P} \alpha_{l,m,p} \psi_{l,p}(n), \tag{15}
\]

\[
z_{u'}^{l}(n) = h_{l,u'}(n) \ast \sum_{m=1}^{M} \sum_{p=1}^{P} \alpha_{l,m,p} \psi_{l,p}(n) \tag{16}
\]

As acknowledged already in \([27], [32], [36], [37]\), the linear and nonlinear signal terms get beamformed towards the same directions. This is clearly visible already in (13) and (14), since the nonlinear basis functions are subject to similar effective beamforming gains of the form \( \sum_{m=1}^{M} e^{j\beta_{l,m,u'}} w_{l,m} \alpha_{l,m,p} \).

Therefore, when multi-beam analog beamformers are adopted in different subarrays, there are as many harmful directions for the distortion, per subarray, as there are intended users. However, very importantly, it can also be observed that apart from the linear filtering effect, the signals in (16) and (18) are both basically identical polynomials of the original digital signal samples \( x_{l}(n) \), expressed through the SNL basis functions \( \psi_{l,p}(n) \) and the effective or equivalent PA coefficients of the whole subarray. Thus, the observable nonlinear distortion at the two considered receivers, contributed by one subarray, is essentially the same, except for the linear filtering, and can be thus modeled with the same polynomial. This implies that a single DPD per subarray can simultaneously provide linearization towards all the intended receivers, which is essential, since the nonlinear distortion from individual subarrays is strongest due to beamforming towards these directions. This forms the technical basis for the proposed DPD system and parameter learning principles described in the next section.

\section{IV. PROPOSED DPD SYSTEM AND PARAMETER LEARNING SOLUTION}

Based on the above nonlinear distortion analysis, we now proceed to formulate the DPD processing methods and parameter learning architecture. We will also explicitly show that the observable distortion can be efficiently suppressed through the adopted DPD processing.

\subsection{A. DPD Processing and Observable Distortion Suppression}

Motivated by (16) and (18), and their generalization to \( U \) users, we argue that a single polynomial DPD can model and suppress the nonlinear distortion stemming from the
corresponding subarray towards all intended receivers. Thus, the core DPD processing in the $l$-th TX path is expressed as

$$\tilde{x}_l(n) = x_l(n) + \sum_{q=3 \text{ odd}}^{Q} \lambda_{l,q}^* \psi_{l,q}(n). \quad (19)$$

where $\psi_{l,q}(n)$, $q = 3, 5, \ldots, Q$ denote the DPD basis functions up to order $Q$, while $\lambda_{l,q}$, $q = 3, 5, \ldots, Q$ denote the corresponding DPD coefficients. We have deliberately excluded processing the amplitude and phase of the linear term in (19), as our main purpose is to suppress the nonlinear distortion while linear response equalization is anyway pursued separately in the RX side. Complex-conjugated DPD coefficients in (19) are adopted only for notational purposes, similar to the classical adaptive filtering literature.

Assuming that the above type of DPD processing is executed in every TX path, we will next explicitly show that the total observable nonlinear distortion can be efficiently suppressed as long as the DPD coefficients are properly optimized. To this end, we substitute the DPD output signals in (19), for $l = 1, 2, \ldots, L$, as the PA input signals in the basis functions in (13), which yields

$$z_u(n) = \sum_{l=1}^{L} h_{l,u}(n) \sum_{m=1}^{M} e^{j\beta_{l,m,u}} \alpha_{l,m,1} w_{l,m} \psi_{l,1}(n)$$

$$+ \sum_{l=1}^{L} h_{l,u}(n) \sum_{m=1}^{M} \sum_{q=3 \text{ odd}}^{Q} e^{j\beta_{l,m,u}} \lambda_{l,q}^* \alpha_{l,m,1} w_{l,m} \psi_{l,q}(n)$$

$$+ \sum_{l=1}^{L} h_{l,u}(n) \sum_{m=1}^{M} \sum_{p=3 \text{ odd}}^{P} e^{j\beta_{l,m,u}} \alpha_{l,m,p} w_{l,m} \psi_{l,p}(n). \quad (20)$$

In above, the first line corresponds to the linear signal while the rest are nonlinear terms. In reaching the above expression it was further assumed that the nonlinear terms introduced by the DPD in (19) are clearly weaker than the linear signal - an assumption that essentially holds in practice - and hence themselves only excite the linear responses of the PAs.

For notational simplicity, we next further assume that the DPD nonlinearity order $Q$ is equal to the PA nonlinearity order $P$, which allows us to rewrite (20) as

$$z_u(n) = \sum_{l=1}^{L} h_{l,u}(n) \sum_{m=1}^{M} \alpha_{l,m,1} e^{j\beta_{l,m,u}} w_{l,m} \psi_{l,1}(n)$$

$$+ \sum_{l=1}^{L} h_{l,u}(n)$$

$$\sum_{m=1}^{M} \sum_{p=3 \text{ odd}}^{P} (\lambda_{l,p}^* \alpha_{l,m,1} + \alpha_{l,m,p}) e^{j\beta_{l,m,u}} w_{l,m} \psi_{l,p}(n). \quad (21)$$

(21) can be re-written as

$$z_u(n) = \sum_{l=1}^{L} h_{l,u}(n) \sum_{m=1}^{M} \alpha_{l,m,1} \psi_{l,1}(n)$$

$$+ \sum_{l=1}^{L} h_{l,u}(n) \sum_{m=1}^{M} \sum_{p=3 \text{ odd}}^{P} (\lambda_{l,p}^* \alpha_{l,m,1} + \alpha_{l,m,p}) u_{l,p}(n). \quad (22)$$

By using the equivalent PA coefficients of the whole subarray, denoted by $\alpha_{l,1}^{\text{tot}} = \sum_{m=1}^{M} \alpha_{l,m,p}$, where the coefficients of the individual $M$ PAs are combined, (22) can be finally expressed as

$$z_u(n) = \sum_{l=1}^{L} h_{l,u}(n) \alpha_{l,1}^{\text{tot}} \psi_{l,1}(n)$$

$$+ \sum_{l=1}^{L} h_{l,u}(n) \sum_{m=1}^{M} \sum_{p=3 \text{ odd}}^{P} (\psi_{l,m}^{\text{tot}} \alpha_{l,m,1}^\text{tot} + \alpha_{l,m,p}^{\text{tot}}) u_{l,p}(n). \quad (23)$$

Based on (23), one can explicitly see that the DPD coefficients $\lambda_{l,q}$ can be chosen such that the nonlinear distortion at the receiver end is suppressed, i.e., $\lambda_{l,q}^{\text{tot}} \alpha_{l,1}^{\text{tot}} + \alpha_{l,q}^{\text{tot}} = 0$. This thus more formally shows that $L$ polynomial DPDs, one per subarray, can effectively linearize $L \times M$ different PAs, particularly when considering the observable linear distortion at RX side, despite all the PA units being generally different. The above expression also shows that despite the observable nonlinear distortion is subject to linear filtering, a memoryless DPD can completely suppress the nonlinear distortion if the PA units themselves are memoryless. Importantly, the expression in (23) also indicates that DPD coefficients that yield good nonlinear distortion suppression are independent of the actual channel realization. Thus, while the beamforming coefficients should obviously follow the changes in the channel characteristics, the DPD system needs to track changes only in the PAs. This will be also verified and demonstrated through the numerical experiments.

Finally, if there is some actual memory in the PA units, the DPD processing in (19) can be generalized such that actual multi-tap digital filters are used instead of scalar coefficients ($\lambda_{l,q}$). In such cases, one can relatively straight-forwardly show that similar conclusions and findings hold as in the memoryless case, i.e., single memory-polynomial DPD unit per TX chain is sufficient for linearization. We provide a concrete numerical example to verify this, in addition to other numerical experiments, in Section V.

**B. Combined Feedback based DPD Learning**

In reality, the nonlinear responses of the individual PA units are unknown and can also change over time. Thus, proper parameter learning is needed. To mimic the over-the-air propagation and thus the true nonlinear distortion at intended receivers, the proposed DPD parameter learning builds on coherently combined observations of the subarray signals. More specifically, as shown already in Fig. 2, the feedback signal in the $l$-th TX path or DPD unit is built by combining the
PA output signals of the corresponding subarray. To this end, and considering the PA output signals in (10), the baseband combined feedback signal in the $l$-th transmitter or subarray reads

$$z_{fb}^l(n) = \sum_{m=1}^{M} w_{l,m}^* y_{l,m}(n)$$  \hspace{1cm} (24)

$$= \sum_{m=1}^{M} |w_{l,m}|^2 \sum_{p=1}^{P} \alpha_{l,m,p} x_l(n) |x_l(n)|^{p-1}$$  \hspace{1cm} (25)

$$= \sum_{m=1}^{M} \sum_{p=1}^{P} \alpha_{l,m,p} x_l(n) |x_l(n)|^{p-1}$$  \hspace{1cm} (26)

$$= \sum_{p=1}^{P} \alpha_{l,p} \psi_l(n).$$  \hspace{1cm} (27)

As can be observed, the combined feedback signal is structurally identical to the actual received signal model in (16), except for the linear filtering effect, forming thus good basis for DPD coefficient optimization.

Generally-speaking the feedback signal model in (27) allows for multiple alternative approaches for DPD parameter learning. One option is to do direct least-squares (LS) based estimation of the effective coefficients $\alpha_{l,p}^{tot}$, and then use these estimates together with (23) to solve for the DPD coefficients $\lambda_{l,p}$ through $\lambda_{l,p}^{tot} \alpha_{l,p}^{tot} + \lambda_{l,p}^{tot} = 0$. Another alternative would be to deploy indirect learning architecture (ILA) [48], [49] where the combined feedback signal in (27) is fed into a polynomial post-distorter whose coefficients are estimated through, e.g., LS, and then substituted as an actual predistorter.

In this article, however, inspired by our earlier work in [36] in the context of single-user MIMO, we pursue closed-loop adaptive learning solutions through the so-called decorrelation principle. Specifically, the DPD learning system seeks to minimize the nonlinear distortion observed at intended users by minimizing the correlation between the nonlinear distortion in the combined feedback signal and the DPD SNL basis functions $\psi_{l,q}(n), q = 3, 5, \ldots Q$. Such learning procedure is carried out in parallel in all $L$ transmitters. To extract the effective nonlinear distortion in the combined feedback signal $z_{fb}^l(n)$, we assume that an estimate of the complex linear gain, denoted by $\hat{G}_l$, is available. Based on this, the effective nonlinear distortion can be extracted as

$$e_l(n) = z_{fb}^l(n) - \hat{G}_l x_l(n).$$  \hspace{1cm} (28)

In practice, $\hat{G}_l$ can be obtained, e.g., by means of block LS. The exact computing algorithm, seeking to tune the DPD coefficients to decorrelate the feedback nonlinear distortion or error signal $e_l(n)$ and the SNL basis functions can build on, e.g., well-known LMS or block-LMS [50] and is not explicitly described for presentation compactness. Additionally, as discussed in [36] in the single-user MIMO context, the SNL basis functions can be mutually orthogonalized through, e.g., QR or Cholesky decompositions, in order to have a faster and smoother convergence.

**V. Numerical Results**

In this section, a quantitative analysis of the performance of the proposed DPD architecture and parameter learning solution is presented by means of comprehensive Matlab simulations.

**A. Evaluation Environment and Assumptions**

The evaluation environment builds on the clustered mmWave channel model described in Subsection II-B, containing $C = 6$ clusters each with $R = 5$ rays. We assume that a LOS component is always available and that the Ricean K-factor is 10 dB. The maximum considered excess delay is 60 ns, a number that is well inline with the assumptions in [51]. We further assume that a hybrid MIMO transmitter simultaneously serves $U = 2$ single-antenna users. The overall transmitter is assumed to contain $L = 2$ TX chains and subarrays, each of them having $M = 16$ antenna elements and the corresponding PA units. Therefore, a total of $M_{TOT} = 32$ antennas and PAs are considered. In each subarray, the antenna spacing is half the wavelength. Furthermore, we evaluate...
the performance of the proposed DPD solution for both the single-beam and multi-beam analog beamformers, discussed in Section II-C, for which example array responses are shown in Fig. 3. Subcarrier-wise digital precoders are always calculated through the ZF approach, as shown in (5), complemented with proper sum-power normalization. Perfect channel state information is assumed to be available at the transmitter. 200 MHz carrier bandwidth is assumed as a representative number in mmWave systems, conforming to 3GPP 5G NR specifications [52] with OFDM subcarrier spacing of 60 kHz, $K_{ACT} = 3168$ active subcarriers and FFT size of $K_{FFT} = 4096$. Finally, the PAPR of the composite multicarrier waveform in each TX chain is limited to $8.3$ dB, through iterative clipping and filtering.

For modeling the individual PA units, measurement data from an actual massive MIMO testbed\(^1\) is used, and memoryless polynomials of order $P = 9$ are identified. Due to hardware constraints, the original PA measurements are carried out for 20 MHz bandwidth while are then resampled to the assumed 200 MHz carrier bandwidth to match the evaluation scenario. Example power spectra of the 32 PA output signals are shown in Fig. 4, where clear differences between the characteristics of the individual PAs can be observed. The passband frequency-selectivity seen in the figure is due to the subcarrier-wise baseband precoder.

As the basic performance metrics, we consider the error vector magnitude (EVM) and adjacent channel leakage ratio (ACLR) to evaluate the inband signal quality as well as the corresponding adjacent channel interference due to spectrum regrowth, respectively, as defined in [52] and [53], and both interpreted for the combined signals. The EVM is defined as

$$EVM_{dB} = \sqrt{P_{\text{error}}/P_{\text{ref}}} \times 100\%,$$

where $P_{\text{error}}$ is the power of the error between the ideal signal samples and the corresponding symbol rate complex samples of the combined array output at the intended receiver direction, both normalized to the same average power, while $P_{\text{ref}}$ is the reference power of the ideal signal. On the other hand, the ACLR is defined as the ratio between the combined powers emitted at the intended channel, $P_{\text{intended}}$, and at the right or left adjacent channels, $P_{\text{adjacent}}$, expressed as

$$ACLR_{dB} = 10 \log_{10} \frac{P_{\text{intended}}}{P_{\text{adjacent}}}.$$

In this work, we always define the intended channel as the bandwidth containing 99% of the total transmitted power in the direction of the intended receiver. The adjacent channel has then the same bandwidth.

In all the following numerical results, the DPD nonlinearity order $Q = 9$ in both ($L = 2$) DPD units. The parameter estimation is carried out with the decorrelation-based approach, implemented in a block-adaptive manner, such that each block contains 100,000 samples and a total of 20 iterations are used. Thus, overall, the DPD parameter estimation utilizes 2,000,000 complex samples. Furthermore, the involved effective linear gains $G_l$, $l = 1, 2$, are estimated through ordinary block least-squares.

### B. DPD Performance at Intended Receivers

First, we evaluate and demonstrate the performance of the proposed DPD structure and parameter learning solution from the two intended receiver directions point of view, assuming the example directions and analog beamforming characteristics as shown in Fig. 3. The 32 PA output signals combine through their respective frequency-selective channels towards the intended receivers, and the corresponding power spectra of the effective combined signals are depicted in Fig. 5, without and with DPD. Furthermore, the multi-beam analog

---

\(^1\)Lund University Massive MIMO testbed, http://www.eit.lth.se/mamitheme
TABLE I
EVM AND ACLR RESULTS

|                  | EVM (%) | ACLR L/R (dB) |
|-----------------|---------|---------------|
| Without DPD at UE1 | 3.17    | 37.89 / 37.76 |
| Without DPD at UE2 | 3.15    | 37.95 / 37.73 |
| With proposed DPD at UE1 | 1.25    | 63.55 / 64.73 |
| With proposed DPD at UE2 | 1.27    | 63.43 / 64.01 |

beamformer approach is considered in this example figure, and therefore both subarrays provide simultaneous beams towards both users. Very similar combined signal spectra are obtained when the single-beam analog beamformer is adopted, and are thus not explicitly shown. Table I shows the corresponding numerical EVM and ACLR values, demonstrating excellent linearization performance at both intended users.

Despite the total combined signal qualities at the intended receivers are very similar for both single-beam and multi-beam analog beamformers, there are fundamental differences in how the DPD processing contributes to suppressing the combined nonlinear distortion in these two cases. To explore this further, we next illustrate the combined received signal spectra at one of the intended users, say UE 2, and deliberately consider the contributions of the two TX subarrays separately. First, when the single-beam analog beamformer is considered, the spectra of the combined subarray signals are shown in Fig. 6, without and with DPD. Now, due to the single-beam analog beamformer, the received signal at UE 2 is largely dominated by subarray 2 while the contribution of subarray 1 is substantially weaker. Hence, as can be observed in the figure, the linearization impact of the DPD unit of subarray 2 is substantial, while it is the combined effect of the array isolation and DPD processing that reduces the OOB emissions stemming from subarray 1. The behaviors of the combined subarray signal spectra at UE 1 are very similar, with the roles of the subarrays interchanged, and are thus omitted.

On the other hand, when the multi-beam analog beamformer is adopted, there is then coherent combining taking place from both subarrays towards the considered UE 2. In this case, the array isolation does not essentially help in controlling the OOB emissions but as shown in Fig. 7, the proposed DPD units can now simultaneously linearize the combined signals of multiple beams. Therefore, the good OOB reduction is solely due to the DPD units. Again, the received spectra at the UE 1 behave very similarly, and are thus omitted.

To provide further insight on the roles of the array isolation and the DPD, we continue to explore the two-user scenario such that the angular separation between the two users is varied. Assuming the beamforming characteristics shown in Fig. 3, with the beam directions controlled according to the user directions, we first place the two intended users very close to each other in the angular domain and configure the analog beams accordingly. Their channel responses are thus very similar, except for the exact phase differences due to the geometry of the environment and scattering. Under these assumptions, highly coherent propagation is expected from both subarrays towards the two intended users regardless of the chosen RF beamforming strategy. Then, the location of one of the intended receivers is kept fixed, while the other one gradually moves along a circular trajectory such that the angular separation is increasing, and beamformers are always adjusted accordingly.

The obtained results in terms of the relative ACLR behavior can be found in Fig. 8 and Fig. 9 when the single-beam and the multi-beam analog beamformers are adopted, respectively, averaged over 100 independent channel realizations for each angular separation value. In the figures, we show separately the behavior of the combined out-of-band emissions due to the two subarrays for the so-called direct links (subarray 1 to UE 1 and subarray 2 to UE 2, averaged across the two users) and the so-called cross-links (subarray 1 to UE 2 and subarray
2 to UE 1, averaged again across the two users). The *Array Isolation* refers to the ratio of the combined OOB emissions of the direct links and those of the cross-links, such that the DPD processing units are deliberately set off. The *DPD Gain*, in turn, refers to the average ACLR improvement obtained by using the proposed DPD units, evaluated separately for the cross-links and the direct links.

In the single-beam beamformer case, as can be observed in Fig. 8, when the users are close in angular domain, the array isolation is naturally small while the DPDs provide good linearization also for the cross-links, both aspects being due to the very high similarity between the array channels of the direct and cross-links. On the other hand, as the angular separation starts to increase, the DPD performance at the cross-links decays while the array isolation increases, but the corresponding total gain stays essentially constant. Then, when the multi-beam analog beamformers are adopted, both users essentially experience coherent propagation from both subarrays. In this case, as expected, the array gain is essentially zero while large DPD gains are systematically available for both the direct and the cross-links independent of the angular separation.

These results show and demonstrate that in the case of *multi-beam analog beamformer*, the DPD units provide simultaneous linearization from each subarray towards all users. Additionally, when the *single-beam analog beamformers* are adopted, the combined effect of array isolation and DPD processing will keep the combined OOB power low. Overall, the results and findings along Figs. 5-9 confirm many of the basic hypotheses made in the previous technical sections. Specifically, the results demonstrate and verify that a single DPD unit can linearize a bank of different PAs when viewed from the combined signal point of view. Additionally, the results verify that the DPD units can provide linearization simultaneously towards multiple directions at which coherent combining is taking place, i.e., when multi-beam analog beamformers are adopted.

### C. DPD Performance in Spatial Domain at Intended and Victim Users

While the above examples demonstrate very high-quality linearization at intended receivers in snap-shot like scenarios, we next pursue evaluating the behavior of the unwanted emissions in the overall spatial domain, i.e., at randomly placed intended and victim users. In these evaluations, we first drop the two intended users at randomly drawn directions and calculate the analog and digital beamformers accordingly. In analog domain, multi-beam approach is utilized. The DPD parameters are calculated as described at the end of Subsection IV-B. Then, while keeping the beamformer and DPD coefficients fixed, we drop 10,000 victim receivers at randomly drawn directions, and evaluate the OOB emissions at all these victim receivers. This is then further iterated over different randomly drawn intended RX directions, such that the beamformer coefficients are recalculated, while also re-executing the DPD parameter learning. Changes in any of the involved array channels do not call for new DPD parameter learning, but it is done here in order to gather statistical information of the parameter learning accuracy. Finally, empirical distributions of the ACLRs at the victim receivers as well as at the intended receivers are evaluated.

The obtained empirical ACLR distributions are shown in Fig. 10. First, the two distributions corresponding to the ACLRs at the intended receivers without and with DPD clearly demonstrate reliable high-quality linearization. Then, the ACLR distribution at victim receivers without any DPD processing clearly indicates that the exact ACLR can vary relatively widely depending on the exact array channel realizations. However, when the DPD units are turned on, large systematic ACLR improvement is obtained with the mini-
Fig. 10. Empirical ACLR distributions at intended and victim users, without and with DPD processing.

Fig. 11. Normalized spectra of the combined received signals at the two intended receivers when 11-th order memory polynomial based PA models with 3 memory taps per nonlinearity order are considered. Also the DPD processing is generalized to account for memory.

mum ACLR realization being ca. 55 dB. These distributions show that overall, systematic and reliable linearization can be provided, at both intended and victim receivers, through the proposed approach.

D. Extension to Memory-based PA Units and DPD Processing

While all previous results and the corresponding technical developments in Sections II and III build on purely memoryless PA models and corresponding memoryless DPD processing, we next demonstrate that the proposed DPD concept can be straightforwardly extended to account for PA memory. First, the same PA measurement data is utilized but now more evolved 11-th order memory polynomials with 3 memory taps per nonlinearity order are considered. These identified memory-based PA models are then taken into use in the evaluations. Additionally, the DPD processing in (19) is also extended such that actual FIR filters are used per nonlinear basis function, instead of simple scalars $\lambda_{l,q}$. Specifically, the DPD order is 11 and 3 memory taps per basis function are adopted. Similar to earlier evaluations, 20 gradient-based block-adaptive learning iterations are used, with 100,000 samples per block.

Assuming the multi-beam analog beamforming approach, and the beampatterns and intended UE directions shown in Fig. 3, the combined received signal spectra without and with DPD processing are depicted in Fig. 11. As can be observed, excellent linearization performance is achieved towards both intended users also when the PA units exhibit memory effects.

VI. Conclusions

In this article, we addressed the power amplifier (PA) nonlinear distortion problem in future array systems, with specific emphasis on multiuser hybrid beamforming based transmitters at mmWaves. First, assuming the generic case of subcarrier-wise multiuser digital precoding and phase-based single-beam or multi-beam analog beamforming in the involved sub-arrays, together with nonlinear and mutually different PA units, the essential signal models were derived describing the combined or observable nonlinear distortion at receiving ends. Then, stemming from the derived signal models, a novel DPD architecture and efficient closed-loop parameter learning solutions were described, allowing to simultaneously linearize the observable signals at all directions where coherent combining takes place. Specifically, it was shown that a single DPD unit is capable of suppressing the unwanted emissions stemming from the corresponding subarray towards all the intended receivers, and thus the composite nonlinear distortion observed at the intended receivers is suppressed by the overall DPD system. Additionally, it was shown that efficient linearization is obtained also from arbitrary victim receivers point of view, stemming from the combined effect of the DPD system and the array isolation/beamforming. Extensive numerical performance examples were provided, with specific focus on timely millimeter wave systems, demonstrating and evidencing the excellent linearization performance of the proposed approach. Finally, the proposed approach was also shown to be applicable in cases where the PA units incorporate substantial memory effects, which is an important practical aspect with wideband mmWave PAs.

References

[1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, “Massive MIMO for next generation wireless systems,” IEEE Communications Magazine, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[2] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, “Five disruptive technology directions for 5G,” IEEE Communications Magazine, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[3] Z. Pi and F. Khan, “An introduction to millimeter-wave mobile broadband systems,” IEEE Communications Magazine, vol. 49, no. 6, pp. 101–107, June 2011.
Tradeoff in Power Amplifier Behavioral Models,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, pp. 1510–1520, June 2010.

[47] F. M. Ghannouchi and O. Hammi, “Behavioral modeling and predistortion,” *IEEE Microwave Magazine*, pp. 52–64, Dec. 2009.

[48] R. N. Braithwaite, “A comparison of indirect learning and closed loop estimators used in digital predistortion of power amplifiers,” in 2015 *IEEE MTT-S International Microwave Symposium*, May 2015, pp. 1–4.

[49] L. Anttila, P. Händel, O. Mylläri, and M. Valkama, “Recursive learning-based joint digital predistorter for power amplifier and I/Q modulator impairments,” *International Journal of Microwave and Wireless Technologies*, vol. 2, no. 2, pp. 173–182, 2010.

[50] S. Haykin, *Adaptive Filter Theory, Fifth Edition*, Pearson, 2014.

[51] 3GPP Tech. Rep. 38.901, “Study on channel model for frequencies from 0.5 to 100 GHz,” v15.0.0 (Release 15), Jun. 2018.

[52] 3GPP Tech. Spec. 38.104, “NR; Base Station (BS) radio transmission and reception,” v15.4.0 (Release 15), Dec. 2018.

[53] 3GPP Tech. Spec. 36.104, “LTE Evolved Universal Terrestrial Radio Access (E-UTRA) Base Station (BS) radio transmission and reception,” v16.0.0 (Release 16), Jan. 2019.