

Digital Predistortion for Multiuser Hybrid MIMO at mmWaves

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

Abstract—Efficient mitigation of power amplifier (PA) nonlinear distortion in multi-user hybrid precoding based broadband mmWave systems is an open research problem. In this article, we carry out detailed signal and distortion modeling in broadband multi-user hybrid MIMO systems, with a bank of nonlinear PAs in each subarray, while also take the inevitable crosstalk between the antenna/PA branches into account. Building on the derived models, we adopt and describe an efficient closed-loop (CL) digital predistortion (DPD) solution that utilizes only a single-input DPD unit per transmit chain or subarray, despite crosstalk, providing thus substantial complexity-benefit compared to the state-of-the-art multi-dimensional DPD solutions. We show that under spatially correlated multipath propagation, each single-input DPD unit can provide linearization towards every intended user, or more generally, towards all spatial directions where coherent propagation is taking place, and that the adopted CL DPD system is robust against crosstalk. Extensive numerical results building on practical measurement-based mmWave PA models are provided, demonstrating and verifying the excellent linearization performance of the overall DPD system in different evaluation scenarios.

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

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]–[5]. Millimeter wave (mmWave) communications 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

This work was financially supported by the Academy of Finland (grants 304147, 301820, 326405 and 319994) and Tampere University Doctoral School.

A. Brihuega, L. Anttila, and M. Valkama are with the Department of Electrical Engineering, Tampere University, Tampere, Finland.

M. Abdelaziz is with Zewail City of Science and Technology, Egypt.

T. Eriksson is with the Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden.

F. Tufvesson is with the Department of Electrical and Information Technology, Lund University, Lund, Sweden.

This paper has supplementary downloadable material available at <http://ieeexplore.ieee.org>, provided by the authors. The material includes measured 28 GHz PA models for two different types of state-of-the-art power amplifiers. This material is 95 kB in size.

bands, and thus large antenna gains are typically needed at both the transmitter (TX) and receiver (RX) ends in order to facilitate reasonable link budgets [1]–[3], [6].

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 [7]. For this reason, many works have proposed and considered hybrid analog-digital beamforming solutions [6]–[14] 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 [3], [12], [14]–[16], which results in reduced multiplexing gain. In general, there are several hybrid architectures depending on how the analog beamforming stage is implemented [6], [7]. 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 a subarray. The reduced complexity architecture is known to be more feasible for practical implementations [6], [7], [11], [17] 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]. 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) [18], leading to nonlinear distortion that degrades the received signal quality of both the in-channel intended receivers as well as the neighboring channel victim receivers. Nonlinear distortion due to PAs in massive MIMO type transmitter systems has been studied in the recent literature [19]–[22]. Specifically, in [19], it was shown that the worst case emissions occur in the direction of the intended receiver, since out-of-band (OOB) distortion also gets beamformed towards this direction.

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) [23], [24]. DPD has been recently studied in the context of large antenna arrays and

MIMO systems in, e.g., [25]–[38]. In [25], [26], [35]–[37], fully digital beamforming based system was investigated. In [25], 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 [26] 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. The MIMO DPD works [35]–[37], in turn, are specifically seeking to tackle the antenna crosstalk problem, which can be considerable especially with closely-spaced antenna elements in large array transmitters [39], [40]. To this end, different indirect learning architecture (ILA) based multi-dimensional DPD solutions have been proposed. However, the amount of the involved basis functions (BFs) and thus the complexity of such multi-input DPD solutions commonly grows steeply with the number of TX paths, being thus prohibitive from the implementation point of view in practical systems, particularly with more than two TX chains.

In [27], [29]–[33], DPD solutions for single-user hybrid MIMO transmitters were investigated assuming the reduced-complexity architecture [7]. 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 [27], the DPD learning is based on observing only a single PA output, within each subarray, while the works in [28], [34] 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, 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 [28].

Other works on hybrid MIMO [30]–[33], [41], [42] 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. In the single-user case, the received signal of the intended user under line-of-sight (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 [30]. The works in [26]–[34], [41], [42] either assume single-user transmission or adopt some other simplifying assumptions such as all PAs being identical, pure LOS propagation or narrowband fading. Thus, nonlinear distortion modeling and corresponding DPD solutions for true multi-user hybrid MIMO systems under mutually different PA units and broadband channels have not been pursued in

the current literature. Moreover, none of the works [25]–[34] tackle the problem of crosstalk, which, on top of the additional implementation related challenges described in [35]–[37], also causes an effective loading phenomenon such that the behavior of the individual PAs are dependent on the array steering angle [39], [41]. This is known as load modulation, and may necessitate continuous learning of the DPD filter coefficients in order to keep track of the fast beam changes [43]. These are the open challenges that this article is addressing.

B. Contributions and Novelty

In this article, we first provide detailed signal and distortion modeling for hybrid-precoded multi-user MIMO-OFDM systems under nonlinear and mutually different PAs affected by input and output crosstalk and containing memory. For generality, the signal models cover both the classical single-beam as well as the more elaborate multi-beam [44] [45] analog beamforming, per subarray. Building on the derived signal and distortion models, we then adopt a closed-loop (CL) single-input DPD solution, structurally similar to that considered in [30] in single phased-array and pure LOS propagation context, while being here deployed in parallel in each TX chain. Due to the considered hybrid precoding based multi-user transmission scenario, the received signals by the intended and potential victim users are contributed by the transmitted signals from all the subarrays, particularly in the multi-beam case but also in the single-beam case due to finite spatial isolation. As a consequence, the overall DPD system needs to provide linearization not only to a single point in space, as was the case in [30], [31], but to multiple points and corresponding receivers – something that is shown to be feasible in this article through the analytical signal models.

For parameter estimation purposes, the PA output signals, per each subarray, are coherently combined in the RF domain, similar to [29], [30], reflecting the radiated nonlinear distortion of each subarray. Importantly, it is analytically shown that such an approach can provide sufficient feedback knowledge for parameter estimation purposes, despite the broadband propagation impacting the true received signals. Additionally, in this article, the inherent robustness of the CL system against antenna crosstalk is established, and leveraged to linearize the transmitter system by means of single-input DPD units. Owing to the CL operating principle and computationally simple learning algorithms, the considered methods may allow for even continuous parameter learning and tracking, which can be essential in mmWave systems with fast beam changes and corresponding changes in the radiated nonlinear distortion [39], [43]. Additionally, compared to the existing multi-dimensional DPD methods and associated extended sets of BFs [35]–[37], the considered single-input DPD units, applied per TX chain or subarray in this work, provide largely reduced complexity and thus more implementation feasible solution in terms of hardware and energy consumption.

Overall, the scope, contributions and novelty of this article can be summarized as follows:

- instead of a single phased-array, linearization of true hybrid MIMO transmitter with multiple simultaneously

operating subarrays with digital precoding and analog beamforming is pursued;

- nonlinear distortion related received signal and system models are derived, incorporating mutually different wideband PA units, crosstalk effects and broadband multipath fading, while also considering both the single-beam and multi-beam analog beamforming cases for generality;
- it is analytically shown that by adopting the single-input closed-loop DPD units, developed in [30] in the single phased-array and LOS propagation context, in all parallel transmitters, nonlinear distortion can be efficiently mitigated despite the strong PA input and PA output crosstalk;
- linearization of a strongly nonlinear multi-user mmWave hybrid array system is successfully shown, through comprehensive numerical examples, containing both single-beam and multi-beam analog beamforming cases, wideband measurement based mmWave PA models, crosstalk and imperfect channel knowledge;
- the results also show that true over-the-air (OTA) feedback is not needed, as the hardware-based subarray-level combiners can provide sufficient feedback knowledge for successful linearization without incorporating the true channel responses.

To the best of the authors' knowledge, this is the first article that successfully shows and demonstrates DPD-based linearization of an mmWave multi-user hybrid beamforming based array system with a bank of nonlinear PA units and broadband propagation.

The remainder of this paper is organized as follows: In Section II, the hybrid multiuser MIMO linear 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 single-input DPD structure and parameter learning solution, while the extension to the challenging crosstalk case is provided in Section V. Section VI provides then a complexity analysis and assessment of the considered CL-DPD and the reference ILA-based DPD. In Section VII, the numerical performance evaluation results are presented and comprehensively analyzed. Lastly, Section VIII will provide the main concluding remarks, while selected details are provided in the Appendix.

II. MULTIUSER HYBRID MIMO - LINEAR SYSTEM MODEL

A. Basics

The overall considered hybrid beamforming based multiuser MIMO-OFDM transmitter is shown in Fig. 1, containing L TX chains and M antenna units per subarray, while serving U single-antenna users simultaneously. The subcarrier-wise BB precoder, denoted as $\mathbf{F}[k] \in \mathbb{C}^{L \times U}$ at subcarrier k , is responsible for mapping the U data streams onto L TX chains and for spatially multiplexing the different users, while the RF precoder, $\mathbf{W} \in \mathbb{C}^{LM \times L}$, focuses the energy towards the dominant directions of the channel. It is further assumed that $U \leq L \leq LM$. The precoded data symbol blocks are then transformed to time-domain OFDM waveforms through IFFTs

of size $K_{\text{FFT}} > K_{\text{ACT}}$ where K_{ACT} denotes the number of active subcarriers. A cyclic prefix is then added to the sample blocks. Finally, it is noted that RX side beamforming can in practice be adopted to improve the received signal strength. However, as the linear signal and the nonlinear distortion exhibit the same beamforming gain, there are no major differences compared to the single-antenna RX case when it comes to assessing the TX-induced nonlinear distortion behavior. Consequently, no RX beamforming is considered in this work for notational simplicity.

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 an extended version of the geometric Saleh-Valenzuela channel model [46], similar to [8]. 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 τ_c , angle of arrival ϕ_c , and angle of departure θ_c , while each ray has its corresponding ray-delay, angle of arrival and angle of departure denoted by τ_r , ν_r and ν_r , respectively. Lastly, let $f_{\text{flt}}(t)$ denote a band-limitation function. Following the above mentioned model, the T_s -spaced delay- d channel vector [8] for the u -th user reads then

$$\mathbf{h}_u[d] = \sum_{c=1}^C \sum_{r=1}^R \rho_{c,r} f_{\text{flt}}(dT_s - \tau_c - \tau_r) a_{\text{Rx}}(\phi_c - \nu_r) \mathbf{a}_{\text{Tx}}(\theta_c - \nu_r) \quad (1)$$

where $\rho_{c,r}$ is the complex gain corresponding to the r -th ray of cluster c , and is drawn from a zero-mean-unit-variance circular symmetric Gaussian distribution, \mathbf{a}_{Tx} denotes the response of the overall TX array, while a_{Rx} accounts for the phase between the clusters and the users. In the special case of uniform linear array (ULA) of size LM , the TX array vector reads $\mathbf{a}_{\text{Tx}}(\zeta) = [1, e^{j2\pi\Delta_{\text{ant}}\sin(\zeta)}, \dots, e^{j(LM-1)2\pi\Delta_{\text{ant}}\sin(\zeta)}]^T$ where Δ_{ant} is the antenna spacing relative to the wavelength.

The corresponding delay- d multiuser MIMO channel matrix reads then $\mathbf{H}[d] = (\mathbf{h}_1[d], \mathbf{h}_2[d], \dots, \mathbf{h}_U[d])^T \in \mathbb{C}^{U \times LM}$. Finally, the corresponding multiuser frequency-domain response at subcarrier k , denoted by $\mathbf{\Gamma}[k] = (\gamma_1[k], \gamma_2[k], \dots, \gamma_U[k])^T \in \mathbb{C}^{U \times LM}$, is given by

$$\mathbf{\Gamma}[k] = \sum_{d=0}^{D-1} \mathbf{H}[d] e^{-j\frac{2\pi kd}{K_{\text{FFT}}}} \quad (2)$$

where D denotes the assumed channel impulse response length in samples. A LOS component can also be added, on top of the channel model in (1), in order to account for Rician fading with any given Rician K-factor defined as the power ratio between the received LOS and NLOS components [47]. Finally, it is noted that a similar channel modeling approach is also utilized in 3GPP mobile radio standardization [48].

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. The analog and digital precoders constitute

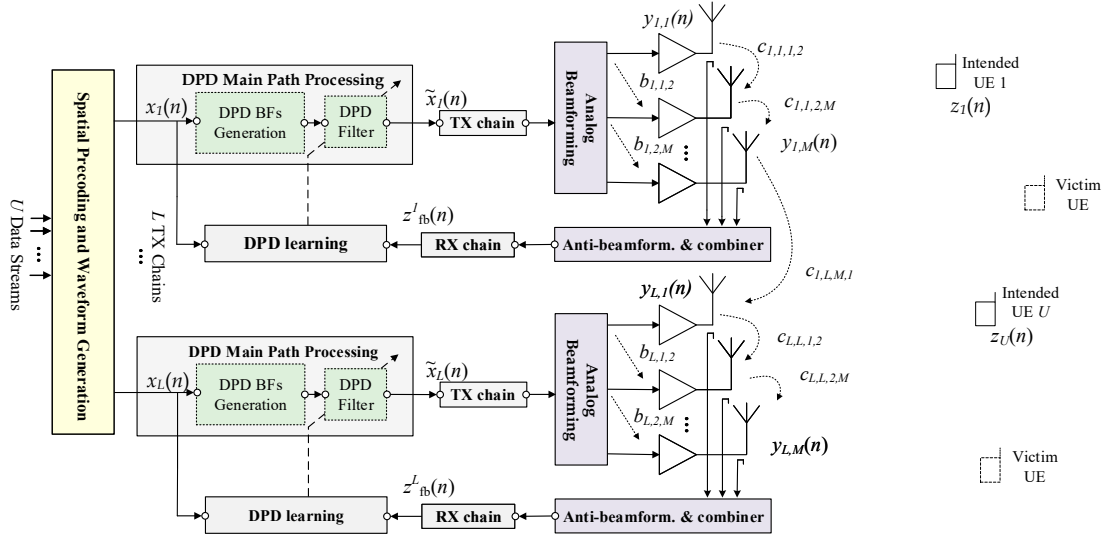


Fig. 1. Block diagram of the considered hybrid beamforming based multiuser MIMO-OFDM transmitter, illustrating also different crosstalk mechanisms at the inputs and outputs of the power amplifiers. Signals $x_l(n)$ and $\tilde{x}_l(n)$ are the input and output of the DPD system in subarray l . Overall, U intended UEs are served, through spatial multiplexing, while also some victim UEs operating in the adjacent channels are shown.

a cascaded system, therefore, both blocks are coupled making the resulting optimization problem non-convex [7]–[9], [12]. 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 [9]. 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 beamformed channel (analog precoder and actual multi-user MIMO channel), is responsible for reducing the multi-user interference and compensating for the frequency-selectivity of the channel. It is noted that the optimization of the precoders is a large topic of its own, while the fairly straight-forward decoupling based approach is considered in this article. Provided that the analog precoder \mathbf{W} 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 $\mathbf{\Gamma}_{\text{eq}}[k] = \mathbf{\Gamma}[k]\mathbf{W}$. For example, the zero-forcing (ZF) and regularized ZF (RZF) precoders essentially read [49], [50]

$$\mathbf{F}_{\text{ZF}}[k] = \mathbf{\Gamma}_{\text{eq}}^H[k](\mathbf{\Gamma}_{\text{eq}}[k]\mathbf{\Gamma}_{\text{eq}}^H[k])^{-1} \quad (3)$$

$$\mathbf{F}_{\text{RZF}}[k] = \mathbf{\Gamma}_{\text{eq}}^H[k](\mathbf{\Gamma}_{\text{eq}}[k]\mathbf{\Gamma}_{\text{eq}}^H[k] + \delta I)^{-1}. \quad (4)$$

For transmit power normalization, additional scaling factors can be introduced, building on, e.g., a sum-power constraint [8], [13], [17]. In practice, the equivalent channels are known only with finite accuracy, an aspect that we address through our numerical experiments in Section VII.

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

$$\mathbf{W} = \mathbf{I}_M \odot (\mathbf{w}_1, \dots, \mathbf{w}_L) \quad (5)$$

where \mathbf{I}_M is an $M \times M$ identity matrix, the operator \odot refers to Khatri-Rao product [51], $\mathbf{w}_l = (w_{l,1}, w_{l,2}, \dots, w_{l,M})^T \in$

$\mathbb{C}^{M \times 1}$ with $|w_{l,m}| = 1 \forall l, m$ is the beamforming vector of the l -th subarray. Interestingly, the weights $w_{l,m}$ can be optimized in multiple ways, while in this work 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 beam patterns 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] and [45].

It is noted that in general, the multi-beam analog beamforming approach can, e.g., facilitate additional macro-diversity, if the different sub-arrays are located over larger distances, while also offering more degrees of freedom for frequency multiplexing. Additionally, as noted in [45], multiple simultaneous beams can also facilitate joint communications and sensing/radar. In this article, we consider both alternative approaches in a comparative manner, while do not explicitly claim one better than the other.

III. MODELING AND ANALYSIS OF PA INDUCED NONLINEAR DISTORTION

To build the basis for linearization, the modeling of the PA-induced nonlinear distortion is next pursued. To account for wideband memory effects in practical mmWave PAs, we consider memory polynomial based PA models in the analysis. Additionally, different PA units are assumed mutually different, no DPD processing is yet considered in this section,

i.e., $\tilde{x}(n) = x(n)$ in the context of Fig. 1, and all modeling is carried out in discrete-time baseband equivalent domain. The extension to the challenging crosstalk case is provided separately, for presentation convenience, in Section V.

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

$$\begin{aligned} y_{l,m}(n) &= \sum_{\substack{p=1 \\ p,\text{odd}}}^P \alpha_{l,m,p}(n) \star v_{l,m}(n) |v_{l,m}(n)|^{p-1} \\ &= w_{l,m} \sum_{\substack{p=1 \\ p,\text{odd}}}^P \alpha_{l,m,p}(n) \star x_l(n) |w_{l,m}x_l(n)|^{p-1}, \end{aligned} \quad (6)$$

where $\alpha_{l,m,p}(n)$ stands for the p th order PA impulse response at the m th antenna branch of the subarray l , while P is the corresponding polynomial order and \star is the discrete-time convolution operator. Since $|w_{l,m}| = 1$, the PA output signal in (6) can be re-written as

$$y_{l,m}(n) = w_{l,m} \sum_{\substack{p=1 \\ p,\text{odd}}}^P \alpha_{l,m,p}(n) \star \psi_{l,p}(n), \quad (7)$$

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 $h_{l,m,u}(n)$ – which corresponds to the sequence of complex numbers at the row of the matrix $(\mathbf{h}_u[0], \dots, \mathbf{h}_u[D-1])$ defined by the considered antenna index – the received signal excluding additive thermal noise for notational simplicity reads

$$z_u(n) = \sum_{l=1}^L \sum_{m=1}^M w_{l,m} h_{l,m,u}(n) \star \sum_{\substack{p=1 \\ p,\text{odd}}}^P \alpha_{l,m,p}(n) \star \psi_{l,p}(n). \quad (8)$$

Assuming next that the individual channels within a single subarray are clearly correlated, a common assumption at mmWaves, one can argue that $h_{l,m,u}(n) \approx h_{l,u}(n)e^{j\beta_{m,u}^l}$, and thus rewrite (8) as

$$z_u(n) \approx \sum_{l=1}^L h_{l,u}(n) \star \sum_{m=1}^M w_{l,m} e^{j\beta_{m,u}^l} \sum_{\substack{p=1 \\ p,\text{odd}}}^P \alpha_{l,m,p}(n) \star \psi_{l,p}(n) \quad (9)$$

where $e^{j\beta_{m,u}^l}$ stems from the phase differences between the signals due to the array geometry as well as exact propagation conditions. Such an approximation is well-argued at mmWaves, where there is typically a dominating LOS path and only few scatterers [12], [17]. 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 [16]. It is also noted that for notational convenience, the phase of the dominant channel tap of $h_{l,u}(n)$

is assumed to be embedded in $e^{j\beta_{m,u}^l}$. Finally, it is noted that no assumptions regarding the correlation of channels between subarrays are made, and that in numerical evaluations in Section VII, array propagation and received signals are modeled and calculated without any approximations.

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

$$z_u^l(n) = h_{l,u}(n) \star \sum_{m=1}^M \sum_{\substack{p=1 \\ p,\text{odd}}}^P e^{j\beta_{m,u}^l} w_{l,m} \alpha_{l,m,p}(n) \star \psi_{l,p}(n) \quad (10)$$

$$z_{u'}^l(n) = h_{l,u'}(n) \star \sum_{m=1}^M \sum_{\substack{p=1 \\ p,\text{odd}}}^P e^{j\beta_{m,u'}^l} w_{l,m} \alpha_{l,m,p}(n) \star \psi_{l,p}(n) \quad (11)$$

It can be seen from (10) and (11) 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 is achieved, implying that $\sum_{m=1}^M e^{j\beta_{m,u}^l} w_{l,m} \alpha_{l,m,p}(n)$ and $\sum_{m=1}^M e^{j\beta_{m,u'}^l} w_{l,m} \alpha_{l,m,p}(n)$ are mutually very similar, for given l and p , up to a possible constant phase difference that is independent of p . More specifically, it can be shown that $\sum_{m=1}^M e^{j\beta_{m,u}^l} w_{l,m} \alpha_{l,m,p}(n) \approx e^{j\xi_{u,u'}^l} (\sum_{m=1}^M e^{j\beta_{m,u'}^l} w_{l,m} \alpha_{l,m,p}(n))$, where the phase $\xi_{u,u'}^l$ does not depend on p . This is because with practical PA implementation technologies and manufacturing processes, it is primarily the implementation inaccuracies that constitute to the mutual differences of the PA samples. Hence, (10) and (11) can be essentially re-written as

$$z_u^l(n) = h_{l,u}^{\text{eff}}(n) \star \sum_{\substack{p=1 \\ p,\text{odd}}}^P \alpha_{l,p}^{\text{tot}}(n) \star \psi_{l,p}(n) \quad (12)$$

$$z_{u'}^l(n) = h_{l,u'}^{\text{eff}}(n) \star \sum_{\substack{p=1 \\ p,\text{odd}}}^P \alpha_{l,p}^{\text{tot}}(n) \star \psi_{l,p}(n), \quad (13)$$

where $\alpha_{l,p}^{\text{tot}}(n) = \sum_{m=1}^M \alpha_{l,m,p}(n)$ stands for the equivalent p -th order PA impulse response of the whole subarray, while the possible additional user-specific constant scalar that is common for all values of p is lumped to the effective channel impulse response $h_{l,u}^{\text{eff}}(n)$. The above approximation holds very accurately with all the measured PA models shared along the article.

As acknowledged already in [19], [20], the linear and nonlinear signal terms get beamformed towards the same directions. This is clearly visible already in (10) and (11), since the nonlinear basis functions are subject to similar effective

beamforming gains of the form $\sum_{m=1}^M e^{j\beta_{m,u}^l} w_{l,m} \alpha_{l,m,p}(n)$. 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 (12) and (13) are both basically identical memory 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 filter 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 channel filtering, and can be thus modeled with the same memory polynomial. This is one of the key results, as it implies that a single DPD per subarray can simultaneously provide linearization towards all intended receivers – something that is essential, since the nonlinear distortion from an individual subarray is strongest towards these directions due to multi-beam beamforming. This modeling and analysis thus form the technical basis for the adoption of the single-input DPD system and parameter learning principles described in the next section.

IV. SINGLE-INPUT DPD SYSTEM AND PARAMETER LEARNING SOLUTION

Based on the above nonlinear distortion analysis, we now proceed by adopting single-input DPD processing, structurally similar to that considered in [30] in single phased-array context, in the different parallel transmit paths, and analytically show that the observable distortion can be suppressed.

A. DPD Processing and Observable Distortion Suppression

Motivated by (12) and (13), and their generalization to U users, we argue that a single memory polynomial DPD can model and suppress the nonlinear distortion stemming from the corresponding subarray towards all intended receivers. Thus, the core DPD engine [30] in the l th TX path is expressed as

$$\tilde{x}_l(n) = x_l(n) + \sum_{\substack{q=3 \\ q,\text{odd}}}^Q \lambda_{l,q}^*(n) \star \psi_{l,q}(n). \quad (14)$$

where $\psi_{l,q}(n)$, $q = 3, 5, \dots, Q$, denote the DPD basis functions up to order Q , while $\lambda_{l,q}(n)$, $q = 3, 5, \dots, Q$ denote the corresponding DPD filter impulse responses. We have deliberately excluded processing the amplitude and phase of the linear term in (14), as our main purpose is to suppress the nonlinear distortion while linear response equalization is anyway pursued separately in the RX side.

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 filters are properly optimized. To this end, we substitute the DPD output signals in (14), for $l = 1, 2, \dots, L$, as the PA input signals in the basis functions

in (12), which together with summing over L yields

$$\begin{aligned} \tilde{z}_u(n) = & \sum_{l=1}^L h_{l,u}^{\text{eff}}(n) \star \alpha_{l,1}^{\text{tot}}(n) \star \psi_{l,1}(n) \\ & + \sum_{l=1}^L h_{l,u}^{\text{eff}}(n) \star \sum_{\substack{q=3 \\ q,\text{odd}}}^Q \lambda_{l,q}^*(n) \star \alpha_{l,1}^{\text{tot}}(n) \star \psi_{l,q}(n) \\ & + \sum_{l=1}^L h_{l,u}^{\text{eff}}(n) \star \sum_{\substack{p=3 \\ p,\text{odd}}}^P \alpha_{l,p}^{\text{tot}}(n) \star \psi_{l,p}(n), \end{aligned} \quad (15)$$

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 (14) 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 (15) as

$$\begin{aligned} \tilde{z}_u(n) = & \sum_{l=1}^L h_{l,u}^{\text{eff}}(n) \star \alpha_{l,1}^{\text{tot}}(n) \star \psi_{l,1}(n) + \sum_{l=1}^L h_{l,u}^{\text{eff}}(n) \\ & \star \sum_{\substack{p=3 \\ p,\text{odd}}}^P (\lambda_{l,p}^*(n) \star \alpha_{l,1}^{\text{tot}}(n) + \alpha_{l,p}^{\text{tot}}(n)) \star \psi_{l,p}(n). \end{aligned} \quad (16)$$

Based on (16), one can explicitly see that the DPD filters $\lambda_{l,p}(n)$ can be chosen such that the nonlinear distortion at the receiver end is suppressed, i.e., $\lambda_{l,p}^*(n) \star \alpha_{l,1}^{\text{tot}}(n) + \alpha_{l,p}^{\text{tot}}(n) = 0$ for the considered range of lags. This thus more formally shows that L memory 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 mutually different. Importantly, the expression in (16) also indicates that DPD filters 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, specifically when there is no substantial crosstalk like assumed in this section. This will be also verified and demonstrated through the numerical experiments, while the challenging crosstalk scenario and its implications are addressed in Section V.

Importantly, the expression in (16) also reveals that the multipath channel filtering is essentially common to all nonlinear distortion terms, stemming from an individual subarray, while the term on the second line is identical to coherent sum of the PA output signals of the corresponding subarray. Thus, instead of seeking to use true OTA signals for DPD learning – an option discussed, e.g., in [39] and the references therein – hardware based combining illustrated in Fig. 1 can reproduce coherent summation of the PA output signals, and thus facilitate efficient DPD learning.

B. Combined Feedback based DPD Learning

In practice, the nonlinear responses of the individual PA units are unknown and can also change over time. Thus, proper parameter learning is needed. Building on the above modeling and adopting a similar approach as done in [30] in the single phased-array context, the considered DPD parameter learning builds on the coherently combined observations of the subarray signals as shown already in Fig. 1. To this end, and considering the PA output signals in (7), the baseband combined feedback signal in the l -th TX or subarray reads

$$z_{\text{fb}}^l(n) = \sum_{m=1}^M w_{l,m}^* y_{l,m}(n) \quad (17)$$

$$= \sum_{m=1}^M |w_{l,m}|^2 \sum_{\substack{p=1 \\ p, \text{odd}}}^P \alpha_{l,m,p}(n) \star \psi_{l,p}(n) \quad (18)$$

$$= \sum_{\substack{p=1 \\ p, \text{odd}}}^P \alpha_{l,p}^{\text{tot}}(n) \star \psi_{l,p}(n). \quad (19)$$

As can be observed, the combined feedback signal is structurally identical to the actual received signal model in (12), except for the channel filtering effect, forming thus good basis for DPD filter optimization.

Regarding the feedback implementation, we note that in time division duplex systems, the receiver-side beamformer of the overall transceiver (e.g. the 5G NR base-station) could potentially be re-purposed for the feedback beamformer task, to alleviate the hardware complexity. We also note that the true OTA combined received signals can, potentially, be used for the parameter learning, as discussed at general level in [39] and references therein. Such an approach would simplify the transmitter system, in terms of the feedback couplers, but would also call for specific OTA measurement receivers and some mechanism to provide the measured I/Q signal samples to the transmitter system. However, the potential load-modulation dependency on the steering angle [39] in practical transmitters would make this method inaccurate or difficult to implement. This is further discussed in the following sections. Additionally, the frequency-selectivity of the broadband channels could potentially hinder the DPD parameter learning, in the sense that the DPD system would interpret the channel memory as part of the PAs' memory. Hence, in this work, we assume the hardware combiner based feedback approach.

Generally-speaking the feedback model in (19) allows for multiple alternative approaches for DPD learning. One option is to do direct least-squares (LS) estimation of the effective impulse responses $\alpha_{l,p}^{\text{tot}}(n)$, and then use these estimates together with (16) to solve for the DPD filters $\lambda_{l,p}(n)$ through $\lambda_{l,p}^*(n) \star \alpha_{l,1}^{\text{tot}}(n) + \alpha_{l,p}^{\text{tot}}(n) = 0$. Another alternative would be to deploy indirect learning architecture (ILA) [23] where the combined feedback signal in (19) is fed into a memory polynomial post-distorter whose coefficients are estimated through, e.g., LS, and then substituted as an actual predistorter.

In this article, inspired by our earlier work in [30] in the context of single-user MIMO, we pursue CL adaptive learning solutions through iterative gradient methods. Such

CL approach allows for implementation-feasible parameter tracking, and is also shown in the next section to be robust against the potential crosstalk. Specifically, the learning system seeks to minimize the observable nonlinear distortion by minimizing the correlation between the nonlinear distortion in the combined feedback signal and the DPD SNL basis functions $\psi_{l,q}(n)$, $q = 3, 5, \dots, Q$, over the considered lags in the DPD filters. Such learning procedure is carried out in parallel in all L transmitters. To extract the effective nonlinear distortion in the combined feedback signal $z_{\text{fb}}^l(n)$, we assume that an estimate of the effective linear impulse response $g_l(n) = \alpha_{l,1}^{\text{tot}}(n)$, denoted by $\hat{g}_l(n)$, is available. Based on this, the effective nonlinear distortion can be extracted as

$$e_l(n) = z_{\text{fb}}^l(n) - \hat{g}_l(n) \star x_l(n). \quad (20)$$

In practice, $\hat{g}_l(n)$ can be obtained, e.g., by means of block LS or any other linear estimator, by using the signal of the l th observation receiver and the corresponding known transmit data $x_l(n)$. Assuming that the feedback combiner weights are well-matched to the corresponding main path RF beamformer, the effective linear loop response $g_l(n)$ is not beam-dependent, and thus can be estimated infrequently.

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 block-LMS [52], and reads

$$\lambda_l(i+1) = \lambda_l(i) - \mu_l \mathbf{R}^{-1} \Psi_l^T(i) \mathbf{e}_l^*(i), \quad (21)$$

where i denotes the learning block index, $\lambda_l(i)$ is the vector containing the l th DPD filter coefficients, μ_l denotes the learning rate, while $\mathbf{R} = \mathbb{E}\{\psi(n)\psi(n)^H\}$ is the covariance matrix of the DPD input vector, which can be precomputed. Furthermore, $\Psi_l(i)$ contains the considered basis function samples properly stacked into a matrix over the learning block size, while $\mathbf{e}_l(i)$ is the corresponding error signal vector. As the basis functions are in practice mutually correlated, the self-orthogonalized version of the classical block-LMS, as shown in (21), is adopted in this work to ensure smooth and fast convergence. Alternatively, one could consider to orthogonalize the basis functions matrix, e.g., through Cholesky decomposition. However, the self-orthogonalizing rule is preferred, complexity-wise, since in such approach the orthogonalization transformation is only executed in the parameter learning loop, while the DPD main path processing can utilize the original basis function samples.

V. EXTENSION TO TRANSMITTER SYSTEMS WITH CROSSTALK

Due to the closely spaced antenna units and high level of integration in practical active antenna array transmitters, crosstalk is inevitable [39], [40]. In this section, we thus provide an important extension of the previous developments to hybrid MIMO transmitter systems affected by such crosstalk. Since the PA units of different subarrays or TX chains are commonly implemented on separate integrated circuits, crosstalk prior to the PAs is assumed to take place only within the individual subarrays, as separate ICs imply higher

physical isolation. On the other hand, antenna crosstalk can basically occur between any two antennas. However, as the crosstalk level decays as a function of the physical distance [21], [40], antenna crosstalk from different subarrays is in general weaker, and is modelled linearly in this work, similar to [35], [36]. It is also noted that while the so-called dual-input model in [21], [38] is a more physically-inspired model of the load modulation phenomenon, the PA input/output crosstalk modeling approach adopted in this work provides, under the above assumptions, the same nonlinear distortion products while is also reflecting the nonlinear distortion beam-dependency.

A. Observable Nonlinear Distortion Under Crosstalk

First, to account for PA input crosstalk [35], [36], an individual PA input signal is re-defined as a linear combination of all the phase-shifted signals of the corresponding sub-array, and thus reads

$$\bar{w}_{l,m}(n) = \left(w_{l,m} + \sum_{\substack{i=1 \\ i \neq m}}^M b_{l,i,m} w_{l,i} \right) x_l(n) \quad (22)$$

$$= \bar{w}_{l,m} x_l(n), \quad (23)$$

where $b_{l,i,m}$ is the complex crosstalk or coupling factor between the i th and the m th PA inputs in subarray l , and $\bar{w}_{l,m} = w_{l,m} + \sum_{\substack{i=1 \\ i \neq m}}^M b_{l,i,m} w_{l,i}$ is the corresponding effective analog beamforming coefficient. In general, it can be noted that $|\bar{w}_{l,m}| \neq 1$, which imposes in general a dependency of the PA nonlinear behaviour on the steering angle, similar to load modulation [39]. Such beam-dependent nonlinear behaviour can potentially call for continuous DPD learning, in real time, in order to adapt to and track the changing nonlinear distortion behavior. This is demonstrated and investigated further in Section VII through numerical examples.

Then, in the presence of antenna coupling, the effective PA output or antenna signal is contributed by the signals leaking from all the other antennas, consequently, (7) is re-written as

$$\begin{aligned} \bar{y}_{l,m}(n) &= \bar{w}_{l,m} \sum_{\substack{p=1 \\ p, \text{odd}}}^P \bar{\alpha}_{l,m,p}(n) \star \psi_{l,p}(n) \\ &+ \sum_{k=1}^L \sum_{\substack{i=1 \\ i \neq m}}^M c_{k,l,m,i} \bar{w}_{k,i} \sum_{\substack{p=1 \\ p, \text{odd}}}^P \bar{\alpha}_{k,i,p}(n) \star \psi_{k,p}(n), \end{aligned} \quad (24)$$

where $\bar{\alpha}_{l,m,p}(n) = |\bar{w}_{l,m}|^{p-1} \alpha_{l,m,p}(n)$ and $c_{k,l,i,m}$ is the antenna complex coupling factor from the i th to the m th antenna between the subarrays k and l .

The corresponding received or observable signal by user u can now be shown to read

$$\begin{aligned} \bar{z}_u(n) &= \sum_{l=1}^L h_{l,u}^{\text{eff}}(n) \star \\ &\left(\sum_{\substack{p=1 \\ p, \text{odd}}}^P \bar{\alpha}_{l,p}^{\text{tot}}(n) \star \psi_{l,p}(n) + \sum_{\substack{k=1 \\ k \neq l}}^L \sum_{\substack{p=1 \\ p, \text{odd}}}^P f_{l,k,p}^{\text{tot}}(n) \star \psi_{k,p}(n) \right) \end{aligned} \quad (25)$$

where $\bar{\alpha}_{l,p}^{\text{tot}}(n) = \sum_{m=1}^M \left(e^{j\Delta\beta_{m,m}^{l,l}} + \sum_{\substack{i=1 \\ i \neq m}}^M e^{j\Delta\beta_{i,m}^{l,m}} b_{i,m} + \sum_{\substack{i=1 \\ i \neq m}}^M c_{l,i,i,m} e^{j\Delta\beta_{i,m}^{k,l}} \right) \bar{\alpha}_{l,m,p}(n)$, $\Delta\beta_{i,m}^{k,l} = \beta_i^k + \bar{\beta}_m^l$, where $\bar{\beta}_m^l$ is the phase of the corresponding analog beamformer coefficient, and $f_{l,k,p}^{\text{tot}}(n) = \sum_{m=1}^M \sum_{\substack{i=1 \\ i \neq m}}^M c_{k,l,i,m} e^{j\Delta\beta_{i,m}^{k,l}} \bar{\alpha}_{k,i,p}(n)$. The detailed analysis steps and derivation of (25), starting from (24), are provided in the Appendix A.

Similarly, the output of the l th observation receiver, given originally in (19), can now be re-expressed under crosstalk as

$$\bar{z}_{\text{fb}}^l(n) = \sum_{\substack{p=1 \\ p, \text{odd}}}^P \bar{\alpha}_{l,p}^{\text{tot}}(n) \star \psi_{l,p}(n) + \sum_{\substack{k=1 \\ k \neq l}}^L \sum_{\substack{p=1 \\ p, \text{odd}}}^P f_{l,k,p}^{\text{tot}}(n) \star \psi_{k,p}(n) \quad (26)$$

which has the same structure as the true received signal except for the actual channel filtering. Additionally, the same derivation steps described in Appendix A for (25) also apply here, to reach the final form in (26).

B. Single-Input Closed-Loop DPD Under Crosstalk

Essentially, the models in (25) and (26) contain L memory polynomials (MPs). Thus, one approach to linearize is to adopt L memory polynomials, in parallel, in each TX branch, which means essentially a multi-input DPD unit. However, as the expression for $f_{k,p}^{\text{tot}}(n), k \neq l$ below (25) shows, the $L - 1$ MP terms do not generally combine coherently towards the receivers. Hence, to emphasize implementation-feasible processing complexity, we still pursue linearizing the transmitter system with the individual single-input MP based DPD units, one per TX branch, as described in (14).

Next, by denoting the linear responses ($p = 1$) as $\bar{\alpha}_{l,1}^{\text{tot}}(n) = g_{l,l}(n)$ and $f_{l,k,1}^{\text{tot}}(n) = g_{k,l}(n), k \neq l$, the observable feedback signal in (26) can be re-written as

$$\bar{z}_{\text{fb}}^l(n) = \sum_{k=1}^L g_{k,l}(n) \star x_k(n) + d_l(n), \quad (27)$$

where $g_{k,l}(n)$ is the effective linear impulse response from the input of the k th TX chain to the output of the l th observation receiver, while the total effective nonlinear distortion is lumped in $d_l(n)$. The error signal for the CL parameter learning system is then given by

$$\bar{e}_l(n) = \bar{z}_{\text{fb}}^l(n) - \sum_{k=1}^L \hat{g}_{k,l}(n) \star x_k(n). \quad (28)$$

where the second-term seeks to suppress the linear contribution of each transmitter from a given observation receiver such that an estimate of the prevailing nonlinear distortion is obtained. The actual learning rule follows then (21). Again, similar to the discussion along (20), the estimates of the linear impulse responses can be obtained through, e.g., linear least-squares or any other linear estimator, by using the known transmit signals and the signals of the observation receivers. It is also noted that the coupling responses $g_{k,l}(n), k \neq l$,

TABLE I
 DPD MAIN PATH PROCESSING COMPLEXITY PER SAMPLE AND TOTAL DPD LEARNING COMPLEXITY OF THE HYBRID MIMO TRANSMITTER. $T = 1$ IF THERE IS NO CROSSTALK, AND $T = L$ IF CROSSTALK IS CONSIDERED.

		CL DPD	ILA DPD
DPD	BF gen.	$L(2N_{\text{IBF}} - 1)$	$L(2N_{\text{IBF}} - 1)$
	Filtering	$L(8N_{\text{BF}} - 8)$	$L(8N_{\text{BF}} - 2)$
Learning	BF gen.	–	$LI_{\text{ILA}}(2N_{\text{IBF}} - 1)N_{\text{ILA}}$
	DPD est.	$LI_{\text{CL}}(8N_{\text{BF}}N_{\text{CL}} + 8N_{\text{BF}}^2) + \left[4LT(N_{\text{CL}} + \frac{D_{\text{lin}}}{3})(D_{\text{lin}})^2 \right] + LT(8D_{\text{lin}} - 2)N_{\text{CL}}I_{\text{CL}} + 2LN_{\text{CL}}I_{\text{CL}}$	$\left[4L(I_{\text{ILA}}(N_{\text{ILA}} + \frac{N_{\text{BE}}}{3})(N_{\text{BF}})^2 \right]$

are, strictly-speaking, beam-dependent since they reflect OTA coupling from k th TX to l th observation RX, and thus stem from non-matched main-path beamformer and feedback anti-beamformer. However, it is straight-forward to show that for given l and k , such beam-dependence corresponds to a complex scalar multiplier that can be calculated based on the corresponding beamformer and anti-beamformer weights. Hence, the actual estimation procedures can be executed based on only the potential time-variance of the true physical coupling features.

Compared to widely-applied indirect learning, which is known to largely suffer from the crosstalk [35]–[37] in single-input DPD context, the CL system seeking to decorrelate the prevailing error signal and the single-input DPD basis functions provides inherent robustness against the crosstalk. This is because in the CL system, the basis functions utilized in the parameter learning are calculated from the clean digital signal samples (i.e., $x_l(n)$) while in ILA-based learning the post-distorter basis functions are calculated using the feedback samples (i.e., $z_{\text{fb}}^l(n)$) which suffer from crosstalk. These are demonstrated through simulations in Section VII-D.

VI. COMPUTATIONAL COMPLEXITY ANALYSIS

In this section, a complexity analysis of the CL DPD solution with self-orthogonalized learning and the widely adopted ILA-based DPD with LS fitting for parameter learning is conducted in the presence and absence of crosstalk. We differentiate between the complexity of the DPD learning, and the complexity of the actual main path linearization. The complexity of the main path is usually more critical as it is to be executed in real time along with the data transmission, while the parameter learning is executed whenever the operation conditions of the transmitter change, e.g., carrier frequency, bandwidth, power and potentially the beam direction. We consider the number of floating point operations per sample (FLOPs) as the complexity metric. For clarity, the notations used in this section are summarized as follows: $\lceil \cdot \rceil$ denotes the ceil operator, N_{IBF} stands for the total number of instantaneous BFs, i.e., the filter lags are excluded from this number, while $N_{\text{BF}} = N_{\text{IBF}}D_{\text{DPD}}$ denotes the total number of BFs, where D_{DPD} refers to the number of taps of an individual DPD filter. Additionally, D_{lin} denotes the length of the linear filter(s) used for linear response modeling in the adopted CL DPD in order to generate the error signal. Furthermore, as earlier, L is the number of subarrays, N_{CL} and N_{ILA} denote the learning block-sizes of the CL DPD and the ILA-based DPD systems,

respectively, while I_{CL} and I_{ILA} are the corresponding numbers of block iterations for the two techniques. Lastly, T is a variable that reads $T = L$ and $T = 1$ when analyzing the CL DPD complexity with and without crosstalk, respectively. This takes into account the difference in the error signal generation complexity in cases without and with crosstalk.

The complexity expressions in their general form are gathered in Table I, while the exact complexity numbers for a specific evaluation scenario are provided along with the numerical results in Section VII. To derive the expressions in Table I, a complex multiplication is assumed to cost 6 FLOPs.

It can be generally noted that inverting strongly nonlinear systems commonly calls for relatively high polynomial orders, especially if very high-quality linearization is pursued. In mmWave active arrays, when emphasizing power-efficiency, the individual PA units are operating in large compression close to saturation [18], [39], and thus the system can be considered strongly nonlinear implying that the level of nonlinear distortion is high. While in our following numerical experiments we utilize memory polynomials of order 11, to demonstrate very high-quality linearization of measured mmWave PA units, it is noted that in practice sufficient linearization may also be obtained with lower polynomial orders, particularly since the linearity requirements at mmWaves [53] are somewhat relaxed compared to lower-frequency systems.

VII. NUMERICAL RESULTS

In this section, a quantitative analysis of the performance of the considered CL DPD architecture and parameter learning solution is pursued and 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 [48]. The maximum considered excess delay is 60 ns, a number that is well inline with the assumptions in [48]. 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 $M = 32$ antenna elements and the corresponding PA units. Therefore, a total of 64 antennas and PAs are considered. In each subarray, the antenna

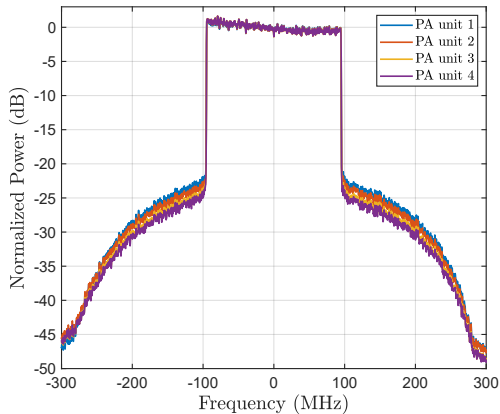


Fig. 2. Example normalized individual PA output spectra of 4 different PA models extracted from measurements on a set of mmWave PAs operating at 28 GHz carrier frequency.

spacing is half the wavelength. It is noted that the developed methods are not limited to any particular values of U , L or M , while the above values reflect an implementation feasible example scenario. Additionally, it is noted that no simplifying assumptions or approximations are taken in the numerical experiments but the received signals are numerically calculated in their exact forms.

Furthermore, we evaluate the performance of the adopted DPD solution for both the single-beam and multi-beam analog beamformers, discussed in Section II-C, though we put most emphasis on multi-beam case as that reflects the true multiuser processing more genuinely. To design the analog beamforming phases, we use the optimization approach described in [44], with angle information as the input. Subcarrier-wise digital precoders are always calculated through the ZF approach, as shown in (3), complemented with proper sum-power normalization. In the basic evaluations, perfect channel state information (CSI) is assumed to be available at the transmitter, in terms of the angles as well as equivalent subcarrier-level complex channel matrices, while the impact of imperfect CSI is addressed in Sub-section VII-D. 200 MHz carrier bandwidth is assumed as a representative number in mmWave systems, conforming to 3GPP 5G NR specifications [53] with OFDM subcarrier spacing of 60 kHz, $K_{ACT} = 3168$ active subcarriers and FFT size of $K_{FFT} = 4096$. Finally, time-domain windowing is adopted to improve the spectral containment of the OFDM signals, while the peak-to-average-power ratio (PAPR) of the composite multicarrier waveform in each TX chain is limited to 8.3 dB, through iterative clipping and filtering (ICF). No beam-tapering is applied in these results, but it is noted that the methods work also with tapering.

For modeling the individual PA units, measurement data from an actual set of mmWave PAs, specifically the PA units in Anokiwave AWMF-0108 ICs, operating at 28.0 GHz carrier frequency is used, and memory polynomials of order $P = 11$ and memory depth $D_{DPD} = 3$ are identified utilizing the same test signal as the one described above. The measured PA models are also shared along the article, together with other sets of PA models obtained by measuring Analog De-

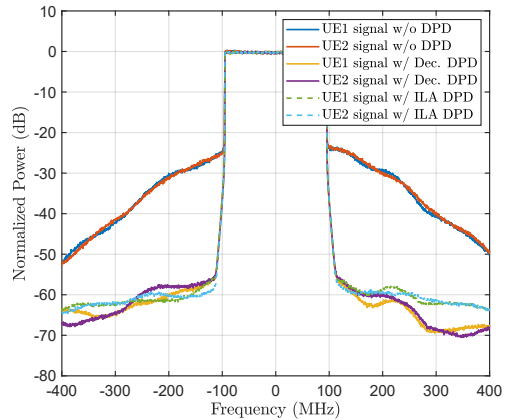


Fig. 3. Normalized combined spectra at the two intended users, without and with DPD, when the *multi-beam analog beamformer* is adopted. For reference and comparison purposes, both the CL DPD and the reference ILA DPD results are shown. No crosstalk is assumed.

vices HMC943APM5E PA ICs at 28.0 GHz and 28.5 GHz. The modeling accuracy of the measured memory polynomial models is in the order of 33 dB in terms of normalized mean squared error (NMSE). Example representative power spectra of four PA output signals are shown in Fig. 2, where substantial nonlinear distortion as well as mutual differences between the characteristics of the individual PA units can be observed.

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 [53] and [54], and both interpreted for the combined signals. The EVM is defined as

$$EVM_{\%} = \sqrt{P_{\text{error}}/P_{\text{ref}}} \times 100\%, \quad (29)$$

where P_{error} is the power of the error between the ideal symbols and the corresponding symbol rate complex samples of the combined array output at the intended receiver direction, after amplitude and phase equalization at subcarrier level, with both signals normalized to the same average power, while P_{ref} is the reference power of the ideal signal. It is noted that the ICF-based PAPR mitigation processing results in an EVM floor of ca. 2% which thus serves as the reference in the evaluations. On the other hand, the ACLR is defined as the ratio between the power observed at the intended channel, P_{intended} , and that at the right or left adjacent channels, P_{adjacent} , interpreted for the effective combined signal at the RX, and expressed as

$$ACLR_{\text{dB}} = 10 \log_{10} \frac{P_{\text{intended}}}{P_{\text{adjacent}}}. \quad (30)$$

It is noted that the ACLR levels in mmWave systems, without linearization, are commonly very low [53], in the order of only 25 dB, reflecting thus very nonlinear system and PA saturation.

In all the following numerical results, the DPD nonlinearity order $Q = 11$ and the memory depth $D_{DPD} = 3$ in both ($L = 2$) DPD units, which results in $N_{\text{IBF}} = 6$ and $N_{\text{BF}} = 18$. The parameter estimation is carried out with the decorrelation-based approach, implemented in a block-adaptive manner, such

TABLE II
EVM AND ACLR RESULTS WITHOUT CROSSTALK AND WHEN
MULTIBEAM ANALOG BEAMFORMING IS ADOPTED

	EVM (%)	ACLR L / R (dB)
Without DPD at UE1	6.41	27.19 / 28.94
Without DPD at UE2	7.42	27.03 / 28.73
With Closed-loop DPD at UE1	2.25	60.41 / 59.73
With Closed-loop DPD at UE2	2.32	59.45 / 58.41
With ILA DPD at UE1	2.31	59.45 / 58.41
With ILA DPD at UE2	2.35	59.61 / 58.78

TABLE III
DPD MAIN PATH PROCESSING COMPLEXITY PER SAMPLE AND TOTAL
DPD LEARNING COMPLEXITY OF THE HYBRID MIMO TRANSMITTER FOR
A TOTAL OF 300,000 SAMPLES

		CL DPD (w/o ; w/ crosstalk)	ILA DPD
DPD (FLOPs/sample)	BF gen.	22	22
	Filtering	272	284
	Total	294	306
Learning (MFLOPs)	BF gen.	—	7.8
	DPD est.	102.31 ; 116.95	777.64
	Total	102.31 ; 116.95	785.44

that each block contains $N_{CL} = 20,000$ samples and a total of $I_{CL} = 15$ iterations are used. Thus, overall, the DPD parameter estimation utilizes 300,000 complex samples. Furthermore, for the error signal generation in parameter learning, the effective linear loop and coupling responses, defined along (20) or (27), are estimated through ordinary block least-squares, with 20,000 samples, prior to the DPD parameter learning. For reference and comparison purposes, also single-input ILA DPD results are commonly shown, building on least-squares parameter learning. The same amount of learning samples as in CL DPD is utilized, in the form of $I_{ILA} = 3$ ILA iterations each containing $N_{ILA} = 100,000$ samples.

B. Results without Crosstalk

1) *DPD Performance at Intended Receivers*: First, we evaluate and demonstrate the performance of the CL DPD structure and parameter learning solution from the two intended receiver directions point of view. The 64 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. 3, without and with DPD. Furthermore, the *multi-beam analog 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 II shows the corresponding numerical EVM and ACLR values, demonstrating excellent linearization performance at both intended users. It can also be observed that very similar linearization performance can be obtained through both the CL DPD and the reference ILA DPD, as the crosstalk aspects are not yet considered. The computational complexities associated to both DPD solutions are shown in Table III, evidencing the low complexity stemming

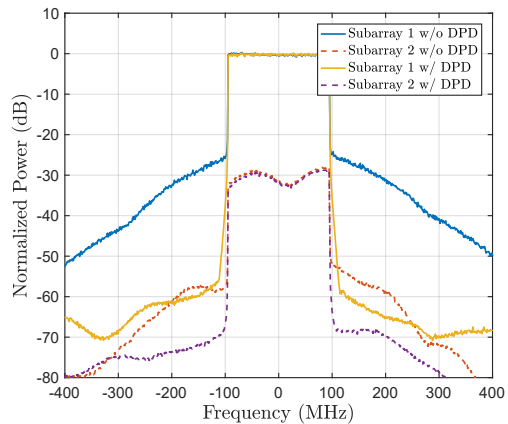


Fig. 4. Normalized spectra of the received combined signals at UE 1, stemming from individual transmit subarrays, considering the *single-beam analog beamformers*. Total received signal is not shown. 'With DPD' refers to the CL DPD method. No crosstalk is assumed.

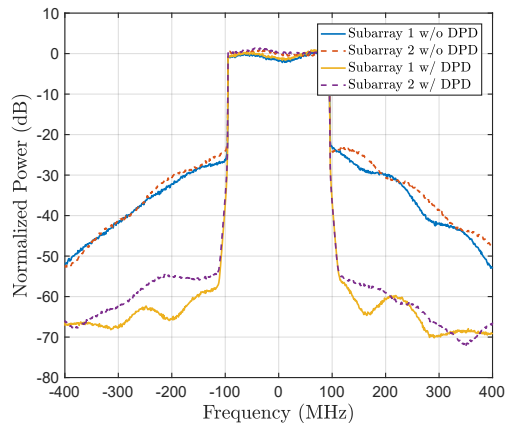


Fig. 5. Normalized spectra of the received combined signals at UE 1, stemming from individual transmit subarrays, considering the *multi-beam analog beamformers*. Total received signal is not shown. 'With DPD' refers to the CL DPD method. No crosstalk is assumed.

from the described CL DPD compared to the widely adopted ILA DPD, which is essential towards real-time implementation and tracking of fast changes in the operation conditions of the array. It is also noted that with lower DPD polynomial orders, the complexity would further reduce, though naturally the linearity would also degrade.

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 1, 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. 4, without and with DPD. Now, due to the single-beam analog beamformer, the received signal at UE 1 is largely dominated by subarray 1 while the contribution of subarray 2 is substantially weaker. Hence, as can be observed in the

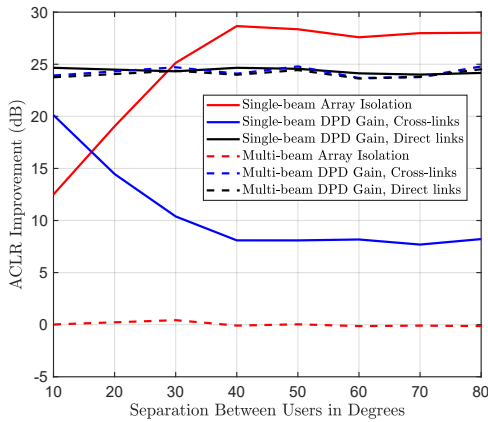


Fig. 6. Impact of the array isolation and the DPD processing on the combined OOB power when the *single-beam analog beamformers* (solid curves) and *multi-beam analog beamformers* (dashed curves) are considered. No crosstalk is assumed.

figure, the linearization impact of the DPD unit of subarray 1 is substantial, while it is the combined effect of the array isolation and DPD processing that reduces the OOB emissions stemming from subarray 2. The behaviors of the combined subarray signal spectra at UE 2 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 1. In this case, the array isolation does not essentially help in controlling the OOB emissions but as shown in Fig. 5, the CL DPD units can now simultaneously linearize the combined signals of multiple beams. Therefore, the good OOB reduction is primarily due to the DPD units. Again, the received spectra at the UE 2 behave very similarly, and are thus omitted.

2) *Impact of DPD gain and array isolation*: 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. 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, while the DPD coefficients are kept fixed for all directions.

The obtained results in terms of the relative ACLR behavior can be found in Fig. 6 when the single-beam and the multi-beam analog beamformers are adopted, averaged over 100 independent channel realizations for each angular separation value. In the figure, 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 crosslinks, when the DPD processing units are deliberately set off. The *DPD Gain*, in turn, refers to the average ACLR improvement obtained by using the CL DPD units, evaluated separately for the cross-links and the direct links.

In the single-beam beamformer case, 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 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. 4-7 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.

3) *DPD Performance in Spatial Domain at Intended and Victim Users*: While the above examples demonstrate very high-quality linearization at intended receivers in snapshot 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 case of the victim receivers, the ACLR is calculated as the ratio between the power radiated at the intended channel in the direction of the intended user, and the OOB power radiated at the adjacent channel in the direction of the victim receiver. Additionally, the OOB power is always measured from the particular neighboring channel – left or right – whichever is worse, such that the most challenging cases are reflected in the analysis. 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 system parameter learning is then executed, and after that, 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

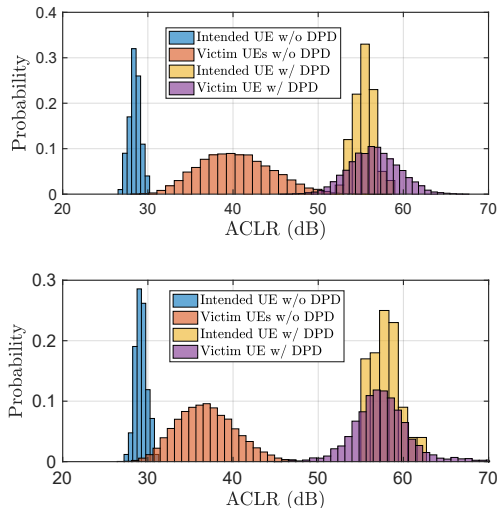


Fig. 7. Empirical ACLR distributions at intended and victim UEs (*i*) when the DPD parameter learning is carried out separately for each random drop of intended UEs (top) and (*ii*) when the DPD parameter learning is executed only for the first drop while fixed for others (bottom). ‘With DPD’ refers to the closed-loop DPD method. No crosstalk is assumed.

beamformer coefficients are recalculated, while experimenting two different scenarios of re-executing and not re-executing the DPD parameter learning. 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. 7, covering the two scenarios in terms of DPD learning. In the first scenario, illustrated on top, the DPD learning is re-executed for every new drop of the intended UE locations, while in the second scenario, illustrated in bottom, the DPD learning is executed only for the first drop of the intended UEs while kept fixed afterwards. Firstly, the ACLR distributions at intended UEs show that the DPD units systematically and reliably improve the signal quality, independent of whether new DPD parameter learning is executed for new array channel realizations or not. Additionally, the ACLR distributions at victim receivers without any DPD processing clearly indicate 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 minimum ACLR realization being ca. 50 dB. Overall, these results show that systematic and reliable linearization can be provided, at both intended and victim receivers, through the adopted approach, and also that changes in the array channels do not call for new DPD parameter learning. As shown in the next subsection, this latter conclusion is, however, valid only when there is no crosstalk.

C. Results with Crosstalk

We next experiment the nonlinear distortion behavior and the CL DPD system linearization performance under crosstalk. As representative baseline values, we set the PA input crosstalk level, within an individual subarray, between two neighboring branches to -20 dB while the corresponding antenna crosstalk level is set to -10 dB [21], [35], [40]. The other coupling

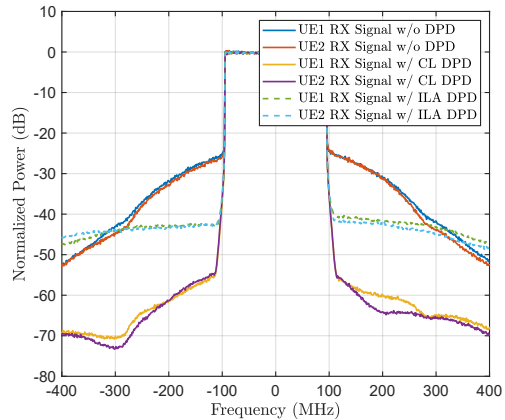


Fig. 8. Normalized combined spectra at the two intended users, without and with DPD, when the *multi-beam analog beamformers* are adopted and with DPD. Baseline PA input crosstalk level is -20 dB while the baseline antenna crosstalk level is -10 dB.

TABLE IV
EVM AND ACLR RESULTS WITH CROSSTALK AND WHEN MULTIBEAM ANALOG BEAMFORMING IS ADOPTED

	EVM (%)	ACLR L/R (dB)
Without DPD at UE1	7.91	29.69 / 29.92
Without DPD at UE2	7.42	29.47 / 29.90
With Closed-loop DPD at UE1	2.89	58.99 / 58.66
With Closed-loop DPD at UE2	3.09	57.32 / 58.21
With ILA DPD at UE1	4.61	40.99 / 41.12
With ILA DPD at UE2	4.55	41.67 / 41.39.

factors between any other two PA input or antenna branches are then assumed to decay as a function of the square of the physical distance and are calculated accordingly. The exact phases of the crosstalk terms, in turn, are always calculated through the effective physical distance between the coupling points, relative to the wavelength. The baseline strength of the crosstalk between two neighboring antenna units is also varied in the evaluations.

An example spectral illustration, under crosstalk and assuming the multi-beam analog beamformer case, is given in Fig. 8, covering also the single-input ILA DPD for reference. As can be observed, the ILA-based reference method is clearly compromised by the crosstalk, while the CL method provides excellent linearization performance despite the relatively strong crosstalk. The corresponding exact EVM and ACLR numbers are shown in Table IV, while the corresponding processing complexities are shown in Table III, again evidencing the lower complexity of the CL DPD solution compared to ILA-based DPD.

To further experiment the impact of the crosstalk on the linearization performance, we next vary the antenna crosstalk level from -15 dB to -10 dB, and evaluate the ACLR performance for both ILA and CL DPD methods. The results are shown in Fig. 9, clearly evidencing the superiority and excellent robustness of the CL DPD under different crosstalk levels. The figure is also showing the corresponding ACLR performance when the subarray size is reduced to $M = 8$. In this case, the superiority of the CL DPD method is even

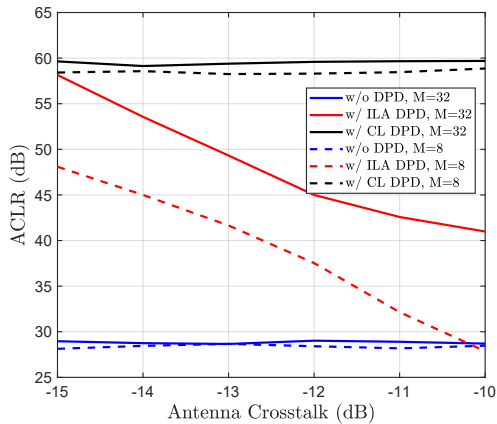


Fig. 9. Obtained ACLR performance as a function of the baseline antenna crosstalk level. The solid lines correspond to the basic evaluation settings (with $M = 32$) while the dashed curves correspond to case with reduced subarray size of $M = 8$.

more obvious, compared to ILA-based DPD, since now there is less help from the noncoherent combining of the crosstalk terms in the combined feedback signal, compared to the case of $M = 32$.

Next, we pursue spatial domain analysis similar to the one in Fig. 7, but now under the baseline crosstalk levels of -20 dB and -10 dB. The obtained ACLR distributions are shown in Fig. 10. As can be observed, when the DPD learning is re-executed for every new intended UE drop (new array channel realization for intended UEs, and hence new analog and digital beamforming coefficients), the adopted CL DPD system provides excellent linearization comparable to the crosstalk-free case in Fig. 7. However, if the DPD coefficients are kept fixed, while the beamforming still follows the changes in the involved channels, a systematic loss in the linearization performance can be observed. This is stemming from the beam-dependency of the nonlinear distortion, caused by the crosstalk [39], [41]. These findings thus imply that continuous or close-to continuous DPD learning is needed, which in turn raises the importance of the computational simplicity of the CL single-input DPD unit and the involved decorrelation based learning. Additionally, the combined feedback observation receiver facilitates the tracking of the changes on the operation conditions due to beamforming. Alternatively, one could consider a predefined set of DPD coefficients for different beam directions [41]. However, adaptability to changes due to, e.g., temperature drifts or device aging would be limited.

D. Results with Crosstalk and Imperfect CSI

Lastly, we address a yet more practical scenario where imperfect CSI is considered in the analog beamforming and digital precoding stages. For the analog beamforming, we still pursue the multibeam scenario through the optimization approach in [44], but then quantize the resulting phase values with 5-bit resolution to yield the actual beamforming weights. For the digital precoding, in turn, we model the estimation errors and uncertainty in the equivalent beamformed channels

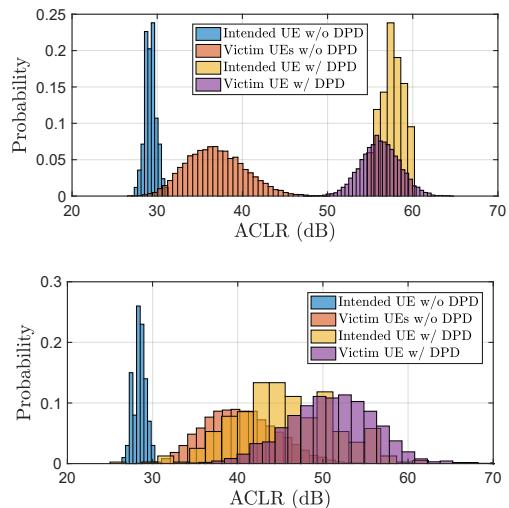


Fig. 10. Empirical ACLR distributions at intended and victim UEs (i) when the DPD parameter learning is carried out separately for each random drop of intended UEs (top) and (ii) when the DPD parameter learning is executing only for the first drop while fixed for others (bottom). 'With DPD' refers to the closed-loop DPD method. Baseline PA input crosstalk level is -20 dB while the baseline antenna crosstalk level is -10 dB.

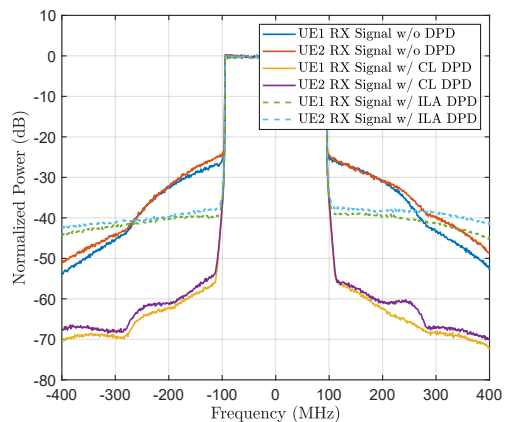


Fig. 11. Normalized combined spectra at the two intended users, without and with DPD, when the *multi-beam analog beamformers* are adopted. Baseline PA input crosstalk level is -20 dB while the baseline antenna crosstalk level is -10 dB. Imperfect CSI in terms of quantized analog beamforming and noisy equivalent channel estimates in digital precoding is considered.

at subcarrier level as [5]

$$\hat{\mathbf{G}}_{\text{eq}}[k] = \chi \mathbf{G}_{\text{eq}}[k] + \sqrt{1 - \chi^2} \mathbf{E}, \quad (31)$$

where $\chi \in [0, 1]$ is a parameter controlling the accuracy of the estimates, while \mathbf{E} is an error matrix whose entries are i.i.d $\mathcal{CN}(0, 1)$ random variables. These estimates are then used to calculate the ZF digital precoder. As a concrete numerical example, the value of $\chi = 0.9$ is used, which corresponds to a power ratio of some ≈ 20 dB between the errors and the true responses. All other parameters, including the crosstalk, are the same as in earlier evaluations.

The obtained results in terms of the example received signal spectra are depicted in Fig. 11. As can be observed, high-quality linearization can still be obtained despite uncertainties in the CSI used for beamforming and precoding. Additionally, the CL DPD is again superior, compared to ILA, primarily

due to the crosstalk.

VIII. 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 with memory, the essential signal models were derived describing the combined or observable nonlinear distortion at receiving ends. Then, stemming from the derived signal models, an efficient single-input DPD architecture and related closed-loop parameter learning solutions were adopted, and shown to allow to simultaneously linearizing the observable signals at all directions where coherent combining takes place. The nonlinear distortion modeling and corresponding DPD developments were then extended to include also the important practical challenge of mutual coupling, such that PA input crosstalk within each subarray as well as generic antenna coupling within the overall large array were both considered.

Specifically, it was shown that the adopted single-input closed-loop 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 the system is very robust against crosstalk and 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 considered approach in different evaluation scenarios involving measured mmWave PA models. Finally, the DPD system was also shown to be applicable and to offer high-performance linearization when the beamforming stages build on imperfect channel state information.

APPENDIX A

DERIVATION OF RECEIVED SIGNAL MODEL (25) UNDER CROSSTALK

In this Appendix, the specific analysis steps to derive the received signal model in (25) are provided. First, by expanding the PA output signal expression in (24) one gets

$$\begin{aligned} \bar{y}_{l,m}(n) &= \bar{w}_{l,m} \sum_{\substack{p=1 \\ p,\text{odd}}}^P \bar{\alpha}_{l,m,p}(n) \star \psi_{l,p}(n) + \sum_{k=1}^L \sum_{\substack{i=1 \\ i \neq m}}^M c_{k,l,m,i} \\ &\times \left(w_{k,i} + \sum_{\substack{i=1 \\ i \neq m}}^M b_{k,i,m} w_{k,i} \right) \sum_{\substack{p=1 \\ p,\text{odd}}}^P \bar{\alpha}_{k,i,p}(n) \star \psi_{k,p}(n) \end{aligned} \quad (32)$$

Since the coupling factors are generally small, commonly around -10 dB to -15 dB [21], [35]–[37], [40] or lower,

the impact of the terms containing products of the form $c_{k,l,m,i} b_{k,i,m}$ can be assumed to be negligible. Then, the combined received signal at the intended user u reads

$$\begin{aligned} \bar{z}_u(n) &= \sum_{l=1}^L \sum_{m=1}^M h_{l,m,u}(n) \star \bar{y}_{l,m}(n) \\ &= \sum_{l=1}^L h_{l,u}(n) \star \\ &\quad \left(\sum_{m=1}^M \left(e^{j\Delta\beta_{m,l}^{l,l}} + \sum_{\substack{i=1 \\ i \neq m}}^M e^{j\Delta\beta_{i,m}^{l,l}} b_{i,m} \right) \sum_{\substack{p=1 \\ p,\text{odd}}}^P \bar{\alpha}_{l,m,p}(n) \star \psi_{l,p}(n) \right) \\ &\quad + \sum_{m=1}^M \sum_{k=1}^L \sum_{\substack{i=1 \\ i \neq m}}^M c_{k,l,i,m} e^{j\Delta\beta_{i,m}^{k,l}} \sum_{\substack{p=1 \\ p,\text{odd}}}^P \bar{\alpha}_{k,i,p}(n) \star \psi_{k,p}(n) \end{aligned} \quad (34)$$

Then, by combining the signal terms in the last line of (34) for $k = l$ with those on the second-last line, and summing over m , we finally reach

$$\begin{aligned} \bar{z}_u(n) &= \sum_{l=1}^L h_{l,u}^{\text{eff}}(n) \star \\ &\quad \left(\sum_{\substack{p=1 \\ p,\text{odd}}}^P \bar{\alpha}_{l,p}^{\text{tot}}(n) \star \psi_{l,p}(n) + \sum_{\substack{k=1 \\ k \neq l}}^L \sum_{\substack{p=1 \\ p,\text{odd}}}^P f_{l,k,p}^{\text{tot}}(n) \star \psi_{k,p}(n) \right), \end{aligned} \quad (35)$$

where $\bar{\alpha}_{l,p}^{\text{tot}}(n)$ and $f_{l,k,p}^{\text{tot}}(n)$ are as defined below (25). This concludes the proof.

REFERENCES

- [1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.
- [2] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, June 2011.
- [3] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!," *IEEE Access*, vol. 1, pp. 335–349, 2013.
- [4] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An Overview of Massive MIMO: Benefits and Challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.
- [5] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.
- [6] J. A. Zhang, X. Huang, V. Dyadyuk, and Y. J. Guo, "Massive hybrid antenna array for millimeter-wave cellular communications," *IEEE Wireless Commun.*, vol. 22, no. 1, pp. 79–87, Feb. 2015.
- [7] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, "Hybrid Beamforming for Massive MIMO: A Survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, 2017.
- [8] A. Alkhateeb and R. W. Heath, "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, May 2016.
- [9] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.
- [10] F. Sotriani and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, April 2016.
- [11] S. Han, C. I. Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.

- [12] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially Sparse Precoding in Millimeter Wave MIMO Systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, March 2014.
- [13] V. Raghavan, S. Subramanian, J. Cezanne, A. Sampath, O. H. Koymen, and J. Li, "Single-User Versus Multi-User Precoding for Millimeter Wave MIMO Systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1387–1401, June 2017.
- [14] M. R. Castellanos, V. Raghavan, J. H. Ryu, O. H. Koymen, J. Li, D. J. Love, and B. Peleato, "Channel-Reconstruction-Based Hybrid Precoding for Millimeter-Wave Multi-User MIMO Systems," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 383–398, 2018.
- [15] P. F. M. Smulders and L. M. Correia, "Characterisation of propagation in 60 GHz radio channels," *Electronics Communication Engineering Journal*, vol. 9, no. 2, pp. 73–80, April 1997.
- [16] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sep. 2015.
- [17] X. Gao, L. Dai, S. Han, C. I, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmwave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, April 2016.
- [18] L. Guan and A. Zhu, "Green communications: Digital predistortion for wideband RF power amplifiers," *IEEE Microw. Mag.*, vol. 15, no. 7, pp. 84–89, Dec. 2014.
- [19] C. Mollen, E. G. Larsson, U. Gustavsson, T. Eriksson, and R. W. Heath, "Out-of-band radiation from large antenna arrays," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 196–203, April 2018.
- [20] L. Anttila, A. Brihuega, and M. Valkama, "On antenna array out-of-band emissions," *IEEE Wireless Commun. Lett.*, vol. 8, no. 6, pp. 1653–1656, Dec 2019.
- [21] K. Hausmair, S. Gustafsson, C. Sanchez-Perez, P. N. Landin, U. Gustavsson, T. Eriksson, and C. Fager, "Prediction of nonlinear distortion in wideband active antenna arrays," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 11, pp. 4550–4563, Nov 2017.
- [22] Y. Zou, O. Raeesi, L. Anttila, A. Hakkarainen, J. Vieira, F. Tufvesson, Q. Cui, and M. Valkama, "Impact of Power Amplifier Nonlinearities in Multi-User Massive MIMO Downlink," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2015, pp. 1–7.
- [23] D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt, and J. Pastalan, "A Generalized Memory Polynomial Model for Digital Predistortion of RF Power Amplifiers," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3852–3860, Oct. 2006.
- [24] H. Jiang and P. A. Wilford, "Digital predistortion for power amplifiers using separable functions," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4121–4130, Aug. 2010.
- [25] M. Abdelaziz, L. Anttila, and M. Valkama, "Reduced-complexity digital predistortion for massive MIMO," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, March 2017, pp. 6478–6482.
- [26] A. Brihuega, L. Anttila, M. Abdelaziz, and M. Valkama, "Digital Predistortion in Large-Array Digital Beamforming Transmitters," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Oct. 2018, pp. 611–618.
- [27] L. Liu, W. Chen, L. Ma, and H. Sun, "Single-PA-feedback digital predistortion for beamforming MIMO transmitter," in *Proc. IEEE Int. Conf. Microw., Millimeter Wave Tech. (ICMMT)*, June 2016, vol. 2, pp. 573–575.
- [28] H. Yan and D. Cabric, "Digital predistortion for hybrid precoding architecture in millimeter-wave massive MIMO systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, March 2017, pp. 3479–3483.
- [29] S. Lee, M. Kim, Y. Sirl, E. R. Jeong, S. Hong, S. Kim, and Y. H. Lee, "Digital Predistortion for Power Amplifiers in Hybrid MIMO Systems with Antenna Subarrays," in *Proc. IEEE Veh. Technol. Conf. (VTC Spring)*, May 2015, pp. 1–5.
- [30] M. Abdelaziz, L. Anttila, A. Brihuega, F. Tufvesson, and M. Valkama, "Digital Predistortion for Hybrid MIMO Transmitters," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 445–454, June 2018.
- [31] X. Liu, Q. Zhang, W. Chen, H. Feng, L. Chen, F. M. Ghannouchi, and Z. Feng, "Beam-Oriented Digital Predistortion for 5G Massive MIMO Hybrid Beamforming Transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 7, pp. 3419–3432, July 2018.
- [32] N. Tervo, J. Aikio, T. Tuovinen, T. Rahkonen, and A. Parssinen, "Digital predistortion of amplitude varying phased array utilising over-the-air combining," in *Proc. IEEE MTT-S Int. Microw. Symposium (IMS)*, June 2017, pp. 1165–1168.
- [33] 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/MTT-S Int. Microw. Symposium (IMS)*, June 2018, pp. 481–484.
- [34] H. Li, G. Li, Y. Zhang, W. Qiao, and F. Liu, "Forward modeling assisted digital predistortion method for hybrid beamforming transmitters with a single PA feedback," in *Proc. IEEE Asia Pacific Conf. Circuits Systems (APCCAS)*, Oct. 2018, pp. 179–182.
- [35] S. Amin, P. N. Landin, P. Händel, and D. Rönnow, "Behavioral Modeling and Linearization of Crosstalk and Memory Effects in RF MIMO Transmitters," *IEEE Trans. Microw. Theory and Techn.*, vol. 62, no. 4, pp. 810–823, April 2014.
- [36] A. Abdelhafiz, L. Behjat, F. M. Ghannouchi, M. Helaoui, and O. Hammi, "A High-Performance Complexity Reduced Behavioral Model and Digital Predistorter for MIMO Systems With Crosstalk," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1996–2004, May 2016.
- [37] Z. A. Khan, E. Zenteno, P. Händel, and M. Isaksson, "Digital Predistortion for Joint Mitigation of I/Q Imbalance and MIMO Power Amplifier Distortion," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 1, pp. 322–333, Jan 2017.
- [38] K. Hausmair, P. N. Landin, U. Gustavsson, C. Fager, and T. Eriksson, "Digital predistortion for multi-antenna transmitters affected by antenna crosstalk," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 3, pp. 1524–1535, March 2018.
- [39] C. Fager, T. Eriksson, F. Barradas, K. Hausmair, T. Cunha, and J. C. Pedro, "Linearity and efficiency in 5G transmitters: New techniques for analyzing efficiency, linearity, and linearization in a 5G active antenna transmitter context," *IEEE Microw. Mag.*, vol. 20, no. 5, pp. 35–49, May 2019.
- [40] M. SalarRahimi, E. Vilela Pinto dos Anjos, P. Taghikhani, V. Volski, C. Fager, D. M. M. Schreurs, G. A. E. Vandenbosch, and M. Geurts, "A Cost-Efficient 28 GHz Integrated Antenna Array With Full Impedance Matrix Characterization for 5G NR," *IEEE Antennas and Wireless Propag. Lett.*, vol. 19, no. 4, pp. 666–670, April 2020.
- [41] E. Ng, Y. Beltagy, G. Scarlato, A. Ben Ayed, P. Mitran, and S. Boumaiza, "Digital Predistortion of Millimeter-Wave RF Beamforming Arrays Using Low Number of Steering Angle-Dependent Coefficient Sets," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 11, pp. 4479–4492, Nov. 2019.
- [42] X. Liu, W. Chen, L. Chen, F. M. Ghannouchi, and Z. Feng, "Linearization for hybrid beamforming array utilizing embedded over-the-air diversity feedbacks," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 12, pp. 5235–5248, Dec 2019.
- [43] C. Fager, K. Hausmair, K. Buisman, K. Andersson, E. Sienkiewicz, and D. Gustafsson, "Analysis of nonlinear distortion in phased array transmitters," in *Proc. Integ. Nonlinear Microw. Millimeter-wave Circuits Workshop (INMMiC)*, Apr. 2017, pp. 1–4.
- [44] M. Mouhamadou, P. Vaudon, and M. Rammal, "Smart Antenna Array Patterns Synthesis: Null Steering and Multi-User Beamforming by Phase Control," *Progress In Electromagnetics Research*, Vol. 60, 95-106, 2006.
- [45] J. A. Zhang, X. Huang, Y. J. Guo, J. Yuan, and R. W. Heath, "Multibeam for joint communication and radar sensing using steerable analog antenna arrays," *IEEE Trans. Veh. Techn.*, vol. 68, no. 1, pp. 671–685, Jan. 2019.
- [46] A. A. M. Saleh and R. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE J. Sel. Areas Commun.*, vol. 5, no. 2, pp. 128–137, Feb. 1987.
- [47] W.C. Jakes, *Microwave Mobile Communications*, IEEE Press, 1974.
- [48] 3GPP Tech. Rep. 38.901, "Study on channel model for frequencies from 0.5 to 100 GHz," v15.0.0 (Release 15), Jun. 2018.
- [49] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.
- [50] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need?," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.
- [51] Khatri, C. G., and Rao, C. Radhakrishna, "Solutions to Some Functional Equations and Their Applications to Characterization of Probability Distributions," *Sankhya: The Indian Journal of Statistics*, vol. 30, no. 2, pp. 167–180, 1968.
- [52] S. Haykin, *Adaptive Filter Theory, Fifth Edition*, Pearson, 2014.
- [53] 3GPP Tech. Spec. 38.104, "NR: Base Station (BS) radio transmission and reception," v15.4.0 (Release 15), Dec. 2018.
- [54] 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.



Alberto Brihuega (S'18) received the B.Sc. and M.Sc. degrees in Telecommunications Engineering from Universidad Politécnica de Madrid, Spain, in 2015 and 2017, respectively. He is currently working towards the Ph.D. degree with Tampere University, Finland, where he is a researcher with the Department of Electrical Engineering. His research interests include statistical and adaptive digital signal processing for compensation of hardware impairments in large-array antenna transceivers.



Fredrik Tufvesson received his Ph.D. in 2000 from Lund University in Sweden. After two years at a startup company, he joined the department of Electrical and Information Technology at Lund University, where he is now professor of radio systems. His main research interests is the interplay between the radio channel and the rest of the communication system with various applications in 5G systems such as massive MIMO, mm wave communication, vehicular communication and radio based positioning.



Lauri Anttila received the M.Sc. and D.Sc. (Hons.) degrees in electrical engineering from Tampere University of Technology (TUT), Tampere, Finland, in 2004 and 2011, respectively. Since 2016, he has been a University Researcher with the Department of Electrical Engineering, Tampere University (formerly TUT). From 2016 to 2017, he was a Visiting Research Fellow with the Department of Electronics and Nanoengineering, Aalto University, Helsinki, Finland. He has coauthored over 100 refereed articles and 3 book chapters. His current research

interests include radio communications and signal processing, with a focus on the radio implementation challenges in systems such as 5G, full-duplex radio, and large-scale antenna systems.



Mahmoud Abdelaziz received the D.Sc. degree (with honors) in Electronics and Communications Engineering from Tampere University of Technology, Finland, in 2017. He then worked as a Post-doctoral Researcher at the same University. He is currently an Assistant Professor at Zewail City of Science and Technology. From 2007 to 2012 he has worked as a communications and signal processing engineer as well as an embedded systems engineer at Newport Media Inc., Etisalat Egypt, and Axxcelera Broadband Wireless. His research interests include

statistical and adaptive signal processing in flexible radio transceivers, in particular, behavioral modelling and digital pre-distortion of power amplifiers in single and multiple antenna transmitters.



Mikko Valkama (S'00–M'01–SM'15) received the M.Sc.(Tech.) and D.Sc.(Tech.) degrees in electrical engineering (EE) from the Tampere University of Technology (TUT), Tampere, Finland, in 2000 and 2001, respectively. In 2003, he was a Visiting Post-Doctoral Research Fellow with the Communications Systems and Signal Processing Institute, San Diego State University (SDSU), San Diego, CA, USA. He is currently a Full Professor and the Department Head of Electrical Engineering with the newly established Tampere University (TAU), Tampere,

Finland. His current research interests include radio communications, radio localization, and radio-based sensing, with particular emphasis on 5G and beyond mobile radio networks. Dr. Valkama was a recipient of the Best Ph.D. Thesis Award of the Finnish Academy of Science and Letters for his Ph.D. dissertation.



Thomas Eriksson received the Ph.D. degree in Information Theory from Chalmers University of Technology, Gothenburg, Sweden, in 1996. From 1990 to 1996, he was at Chalmers. In 1997 and 1998, he was at AT&T Labs - Research, Murray Hill, NJ, USA. In 1998 and 1999, he was at Ericsson Radio Systems AB, Kista, Sweden. Since 1999, he has been with Chalmers University, where he is currently a professor of communication systems. Further, he was a guest professor with Yonsei University, S. Korea, in 2003-2004. He has authored

or co-authored more than 250 journal and conference papers, and holds 14 patents. Prof. Eriksson is leading the research on hardware-constrained communications with Chalmers University of Technology. His research interests include communication, data compression, and modeling and compensation of non-ideal hardware components (e.g. amplifiers, oscillators, and modulators in communication transmitters and receivers, including massive MIMO). Currently, he is leading several projects on e.g. 1) massive MIMO communications with imperfect hardware, 2) MIMO communication taken to its limits: 100Gbit/s link demonstration, 3) wideband linearization (DPD), 4), Satellite communication with phase noise limitations, 5) Efficient and linear transceivers, etc. He is currently the Vice Head of the Department of Signals and Systems with Chalmers University of Technology, where he is responsible for undergraduate and masters education.