

# Beam-Level Frequency-Domain Digital Predistortion for OFDM Massive MIMO Transmitters

Alberto Brihuega<sup>1</sup>, Lauri Anttila<sup>2</sup>, and Mikko Valkama<sup>3</sup>, *Fellow, IEEE*

**Abstract**—In this article, a novel digital predistortion (DPD) solution for fully digital multiple-input–multiple-output (MIMO) transmitters (TXs) is proposed. Opposed to classical DPD solutions that operate at TX chain or antenna level, the proposed DPD operates at the stream or beam level, and hence, its complexity is proportional to the number of spatially multiplexed streams or users rather than to the number of antennas. In addition, the proposed beam-level DPD operates in the frequency domain (FD), which makes it possible to provide flexible frequency-dependent linearization of the transmit waveforms. This feature is very well suited to the linearity requirements applicable at the 5G new radio (NR) frequency-range 2 (FR2), where the inband quality requirements commonly limit the feasible TX power and can also vary significantly within the channel bandwidth depending on the utilized data modulation and coding schemes of the different frequency-multiplexed users. Altogether, the proposed solution enables a large reduction in the computational complexity of the overall DPD system, and its performance is demonstrated and verified through both experimental and simulation-based results.

**Index Terms**—Array transmitters (TXs), beamforming, digital predistortion (DPD), 5G, frequency domain (FD), massive multiple-input–multiple-output (MIMO), millimeter-wave (mmWave) communications, multiuser MIMO, nonlinear distortion, orthogonal frequency-division multiplexing (OFDM), power amplifiers (PAs).

## I. INTRODUCTION

**L**INEARIZATION of active array radio transmitters (TXs) with large numbers of antennas and power amplifier (PA) units is one timely topic in digital predistortion (DPD) research, posing many new challenges compared to classical single PA linearization [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14]. The operation principle and specific implementation of DPD solutions tailored for antenna arrays depend heavily on the considered TX architecture. For instance, in phased arrays or analog-beamforming-based TXs, a single DPD unit is responsible for linearizing a bank of PAs. The common approach is to leverage the spatial characteristics of the radiated nonlinear distortion, which is

most dominant in the direction of the main beam [15], [16]. Hence, the DPD problem builds on minimizing the nonlinear distortion from the beamformed channel perspective, which eventually resembles a classical single-input–single-output (SISO) DPD task [1], [3], [17]. This approach results in the DPD minimizing the distortion around the main beam, while the joint effect of the beamforming and the DPD ensures a sufficiently low level of unwanted emissions elsewhere [1], [2]. A similar approach can be applied to a more general hybrid-beamforming configuration, in which multiple phased arrays, commonly referred to as subarrays, are deployed within the TX, together with an additional digital beamforming or precoding stage. In such hybrid TX architecture, each DPD is responsible for minimizing the far-field beamformed distortion stemming from its corresponding subarray [18]. The linearization principles of such analog and hybrid TX architectures are already relatively well established and well understood. However, less emphasis has been put on finding new approaches to efficiently linearize fully digital array TXs, in which many challenges still remain unsolved—computational complexity being one of the main concerns. This is the main focus area of this article.

In fully digital multiple-input–multiple-output (MIMO) TXs, it is possible to digitally control the input signal to every individual PA, and thus, a DPD per TX/antenna is traditionally considered, as shown in Fig. 1(a). This configuration is equivalent to having multiple parallel single PA linearization blocks, resulting in the complexity growing linearly with the number of antennas. Such architecture is referred to as parallel DPD or per-TX DPD hereafter in this article. A few works have investigated the behavior of nonlinear distortion in fully digital MIMO radios. In particular, Braithwaite et al. [12], Mollén et al. [15], and Anttila et al. [16] investigated what kind of intermodulation distortion (IMD) beams are generated when multiuser spatial precoding is considered in the TX. The work in [12] further showed that it is also possible to cancel such IMD beams. However, most of the existing works tailored for DPD in digital MIMO TXs focus on reducing the learning and running complexity of the DPD units [19], [20], [21], [22], [23] or on developing complex behavioral models capable of linearizing the TX in the presence of potential crosstalk [8], [9], [24], [25]—all inherently assuming the parallel DPD configuration. However, with such parallel DPD approach, the benefits of deploying DPD in fully digital TXs can be limited or challenging to achieve. Specifically, in modern 5G new radio (NR) and beyond MIMO radios, the power consumption needed for linearizing the TX is a great concern, as the

Manuscript received 11 July 2022; revised 23 September 2022; accepted 19 October 2022. This work was supported in part by the Academy of Finland under Grant 304147, Grant 301820, Grant 326405, and Grant 319994; and in part by the Tampere University Doctoral School. Limited set of early-stage results presented at IEEE IMS 2021 [DOI: 10.1109/LMWC.2021.3062982]. (*Corresponding author: Alberto Brihuega.*)

Alberto Brihuega is with Nokia Mobile Networks, 90620 Oulu, Finland (e-mail: alberto.brihuegagarcia@tuni.fi).

Lauri Anttila and Mikko Valkama are with the Department of Electrical Engineering, Tampere University, 33100 Tampere, Finland.

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TMTT.2022.3222320>.

Digital Object Identifier 10.1109/TMTT.2022.3222320

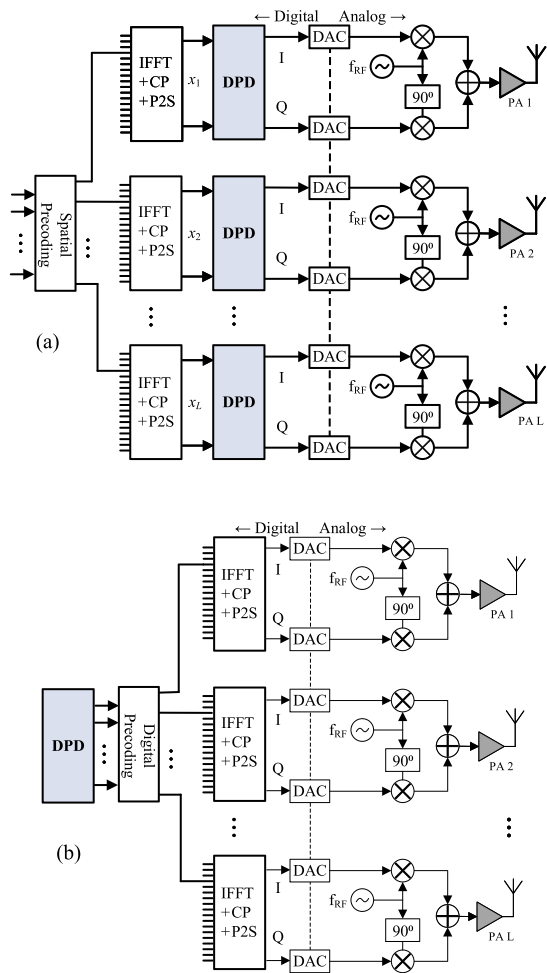


Fig. 1. Conceptual illustration of fully digital MIMO TX system with (a) classical antenna-level or per-TX DPD configurations and (b) proposed stream-/beam-level DPD configuration.

high-power PAs commonly found in legacy base stations are replaced by multiple low-/medium-power units in an antenna array configuration. This makes the deployment of a DPD per PA/antenna very challenging from the energy efficiency point of view and potentially discourages its adoption due to the limited power headroom for DPD operation. In addition, at the millimeter-wave (mmWave) frequency bands or the so-called frequency-range 2 (FR2) in the 5G NR terminology, the instantaneous modulation bandwidth (BW) can be as high as 400 MHz (in Rel-15 and Rel-16) or even 2 GHz in the latest Rel-17, which together with common  $5\times$  oversampling factors for DPD operation make the running complexity of the parallel DPDs practically prohibitive.

In this article, building on our initial work in [11] in the context of single-user phased-array TXs, we propose a new reduced-complexity frequency-domain (FD) DPD architecture for orthogonal frequency-division multiplexing (OFDM)-based digital MIMO TXs, in which the DPD is moved from the antenna-level digital front end to baseband (BB), prior to the digital precoding stage, as shown conceptually in Fig. 1(b). This has several important implications. First, the DPD operates at the stream/beam level, and hence, its complexity grows

only proportionally to the number of dominant spatially multiplexed streams or users, rather than to the number of antennas. Second, this new architecture requires some modifications to the BB operation in classical radios, particularly to the size of the inverse fast Fourier transform (IFFT) and spatial precoding processing. Third, as the DPD runs prior to the OFDM waveform generation stage, it operates in the FD, which makes it possible to easily control the BW of the predistorted signal and to provide flexible frequency-selective linearization of the wideband transmit waveform [11]. These characteristics can be of special interest for systems operating at the FR2 bands, in which the nonlinear operation conditions of the TX are easily limited by the inband quality requirements [2], as the adjacent channel emission limits have been substantially relaxed [26]. Hence, due to the possibility of controlling the DPD linearization in a frequency-dependent manner, it is possible to adapt the linearization performance within the TX passband so that sufficient linearization is provided to fulfill the quality imposed by the different data modulation and coding schemes of the frequency-multiplexed users.

The main contributions of this article can be described and summarized as follows.

- 1) We propose a novel reduced-complexity beam-level DPD architecture that avoids adopting per-TX DPD units, thus largely alleviating the overall DPD system complexity and thereon facilitating efficient DPD of fully digital MIMO TXs.
- 2) We introduce and describe the fundamental models and theories, with specific emphasis on the processing stages required to provide stream-/beam-level linearization in multiuser MIMO scenarios.
- 3) We analyze and exploit the spatial characteristics of the nonlinear distortion in digital MIMO TXs, which are distinct to what are commonly encountered in the context of phased-array TXs, in order to further improve the processing efficiency of the proposed architecture.
- 4) It is shown that the parameter estimation method of the proposed DPD solution is resilient to potential crosstalk, being thus capable of offering superior performance compared to the reference DPD configurations—unless multidimensional basis functions (BFs) were adopted to cope with the crosstalk, which in turn would come with a large increase in computational complexity.
- 5) We provide an explicit complexity analysis of the proposed DPD methods while also comparing against the parallel per-TX DPD approach and also show the complexity tradeoffs as the function of the number of TX paths or antennas.
- 6) Extensive experimental and simulation-based numerical results are provided, covering both phased-arrays and digital MIMO TXs, in order to demonstrate and verify the operation of the proposed DPD methods in different alternative use cases.

The rest of this article is organized as follows. Section II shortly states the multiuser MIMO linear system model and the corresponding assumptions. In Section III, the proposed beam-level DPD concept is introduced, focusing first on the basic single-user case, whereas the extension to the more

general multiuser scenario is then provided in Section IV. The implications of potential crosstalk are addressed and discussed in Section V, while a complexity analysis of the proposed DPD solution and a corresponding complexity comparison against the parallel DPD architecture are provided in Section VI. Extensive experimental and simulation-based results and their analysis are provided in Section VII. Finally, Section VIII provides the final concluding remarks.

## II. LINEAR MIMO-OFDM SYSTEM MODEL

We consider a MIMO-OFDM TX with  $L$  TX chains that simultaneously serves  $U < L$  single-antenna users through spatial multiplexing. The subcarrierwise BB digital precoder is denoted as  $\mathbf{F}[k] \in \mathbb{C}^{L \times U}$ , which maps the  $U$  data streams onto the  $L$  TX chains and multiplexes the users in the spatial domain. Within an individual OFDM symbol, the corresponding output of the BB precoder at subcarrier  $k$  is given by  $\mathbf{x}[k] = \mathbf{F}[k]\mathbf{d}[k]$ , where  $\mathbf{d}[k] \in \mathbb{C}^{U \times 1}$  is the vector containing the spatially multiplexed user data symbols. Then, the precoded subcarrier samples are transformed to time-domain OFDM waveforms through  $K$ -point inverse discrete Fourier transforms (IDFTs) in each of the  $L$  TXs. The corresponding time-domain OFDM waveform samples at the  $l$ th TX read then

$$x_l(n) = \frac{1}{\sqrt{K_{\text{act}}}} \sum_{k=-K_{\text{act}}/2}^{K_{\text{act}}/2-1} x_l[k] e^{j2\pi kn/K_{\text{IDFT}}} \quad (1)$$

where  $K_{\text{act}}$  denotes the number of active subcarriers, with  $K_{\text{IDFT}} > K_{\text{act}}$ , and  $x_l[k]$  is the  $l$ th element in  $\mathbf{x}[k]$ .

In order to model the wideband propagation conditions and the spatial correlation characteristics between the different antenna channels common at mmWaves, an extended version of the Saleh–Valenzuela geometric channel model [27] is adopted. It is assumed that the propagation environment consists of  $C$  clusters, each of them containing  $R$  rays. Each cluster  $c$  has a certain path delay  $\tau_c$ , angle of arrival  $\phi_c$ , and angle of departure  $\theta_c$ , while each ray has its corresponding ray delay, angle of arrival, and angle of departure denoted by  $\tau_r$ ,  $\nu_r$ , and  $\nu_r$ , respectively. The  $T_s$ -spaced channel vector for the  $u$ th user then reads [18], [28]

$$\mathbf{h}_u(d) = \sum_{c=1}^C \sum_{r=1}^R \rho_{c,r} f_{\text{fit}}(dT_s - \tau_c - \tau_r) \times a_{\text{Rx}}(\phi_c - \nu_r) \mathbf{a}_{\text{Tx}}(\theta_c - \nu_r) \quad (2)$$

where  $\rho_{c,r}$  is the complex gain corresponding to the  $r$ th ray of cluster  $c$  and follows a zero-mean, unit-variance circular symmetric Gaussian distribution,  $\mathbf{a}_{\text{Tx}}(\cdot)$  denotes the response of the overall TX array, and  $f_{\text{fit}}(\cdot)$  denotes a band-limitation function.

The corresponding multiuser MIMO channel matrix is denoted by  $\mathbf{H}(d) = (\mathbf{h}_1(d), \mathbf{h}_2(d), \dots, \mathbf{h}_U(d))^T$ , while its FD equivalent is given by

$$\mathbf{\Gamma}[k] = \sum_{d=0}^{D-1} \mathbf{H}[d] e^{-j2\pi kd/K_{\text{IDFT}}} \quad (3)$$

where  $D$  is the maximum length of the channel in samples. The impulse response between the  $l$ th TX antenna and the  $u$ th user is denoted as  $h_{l,u}(n)$ . Due to the high correlation between antenna channels present at mmWaves, the different antenna channels for user  $u$  can be expressed or approximated as [18]

$$h_{l,u}(n) \approx h_u(n) e^{j\beta_l u}, \quad \text{for } l = 1, 2, \dots, L. \quad (4)$$

The above refers to a channel structure containing a common impulse response and antenna-specific complex exponentials that contain information about the phase differences between the antenna signals due to the geometry of the array and the exact propagation conditions. In addition, the phase of the dominant tap of the channel impulse response is assumed to be embedded in  $e^{j\beta_l u}$ . It is noted that under pure line-of-sight (LOS) propagation conditions, the common impulse response is of the form  $h_u(n) = \delta(n)$ .

Provided that the multiuser channel matrix is known and that the simple maximum-ratio-transmission (MRT) principle is adopted, the BB precoder can be obtained as  $\mathbf{F}[k] = \mathbf{\Gamma}^H[k]$ , where the BB precoder is applicable at the passband frequency bins, i.e., for  $-K_{\text{act}}/2 \leq k \leq K_{\text{act}}/2 - 1$ . Finally, the corresponding equivalent BB precoder in the time-domain can be represented in a similar fashion to the channel impulse responses, which for the  $l$ th element and  $u$ th user reads  $F_{l,u}(n) = F_u^{\text{eff}}(n) e^{-j\beta_l u}$ , where  $F_u^{\text{eff}}(n)$  is the common component of the time-domain equivalent BB precoder. Such relation is utilized next, in the analysis of the PA-induced nonlinear distortion in the considered MIMO system, as well as in further restructuring of the FD BB precoder and the actual DPD system.

## III. PROPOSED SINGLE-USER BEAM-LEVEL DPD

In order to establish the basic intuition behind the proposed DPD solutions, the more simple single-user case with  $U = 1$  is considered first, while the extension to the actual multiuser case is provided in Section IV. We start by describing the digitally beamformed PA output signals and the corresponding over-the-air (OTA) combined received signal, under the assumption of memory polynomial (MP) [29] PA models, and then formulate the actual FD DPD processing and the corresponding parameter learning solutions.

### A. Nonlinear System Model and Beamformed Distortion

Given the stimulus waveform  $x_l(n)$  in (1), the  $l$ th PA output signal under the assumption of an MP PA model can be expressed as

$$\begin{aligned} y_l(n) &= \sum_{p=1}^P \sum_{m=0}^M \alpha_{p,m}^l x_l(n-m) |x_l(n-m)|^{2(p-1)} \\ &= \sum_{p=1}^P \sum_{m=0}^M \alpha_{p,m}^l (d(n-m) \star F_l(n-m)) \\ &\quad \times |d(n-m) \star F_l(n-m)|^{2(p-1)} \\ &= \sum_{p=1}^P \sum_{m=0}^M \alpha_{p,m}^l \left( d(n-m) \star F^{\text{eff}}(n-m) e^{-j\beta_l} \right) \end{aligned}$$

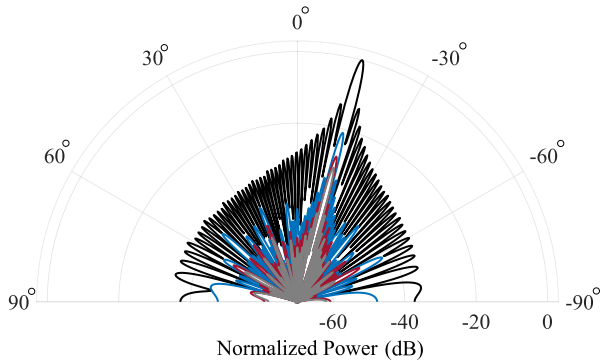


Fig. 2. Example LOS far-field beampatterns for total radiated signal (black), total distortion (blue), passband nonlinear distortion (red), and OOB distortion (gray); 100 antennas and PAs are considered.

$$\begin{aligned}
& \times \left| d(n-m) \star F^{\text{eff}}(n-m) e^{-j\beta_l} \right|^{2(p-1)} \\
& = \sum_{p=1}^P \sum_{m=0}^M \alpha_{p,m}^l \left( d(n-m) \star F^{\text{eff}}(n-m) e^{-j\beta_l} \right) \\
& \quad \times \left| d(n-m) \star F^{\text{eff}}(n-m) \right|^{2(p-1)} \\
& = \sum_{p=1}^P \sum_{m=0}^M \alpha_{p,m}^l e^{-j\beta_l} \Upsilon_{p,m}(n) \quad (5)
\end{aligned}$$

where  $\alpha_{p,m}^l$  refer to the MP model coefficients of the  $l$ th PA,  $\star$  denotes the linear convolution operator, and  $\Upsilon_{p,m}(n) = (d(n-m) \star F^{\text{eff}}(n-m)) |d(n-m) \star F^{\text{eff}}(n-m)|^{p-1}$  is the corresponding  $p$ th-order BF with delay  $m$ .

Then, the  $L$  PA output signals propagate OTA and combine at the receiver end. This is expressed as

$$\begin{aligned}
z_u(n) &= \sum_{l=1}^L h_{l,u}(n) \star y_l(n) \\
&= \sum_{l=1}^L \sum_{p=1}^P \sum_{m=0}^M \alpha_{p,m}^l e^{-j\beta_l} h_{l,u}(n) \star \Upsilon_{p,m}(n) \\
&= h_u(n) \star \sum_{p=1}^P \sum_{m=0}^M \alpha_{p,m}^{\text{tot}} \Upsilon_{p,m}(n) \quad (6)
\end{aligned}$$

where the expression in (4) has been utilized to arrive at the final form,  $\alpha_{p,m}^{\text{tot}} = \sum_{l=1}^L \alpha_{p,m}^l$  denotes the effective PA coefficient from the beamformed channel perspective, and  $h_u(n)$  refers to the common impulse response of the antenna channels. From (6), it is possible to observe that the nonlinear distortion is beamformed also toward the direction of the intended receiver, while in other directions, it gets diluted due to noncoherent propagation. This is shown in Fig. 2 in the simple example case of LOS propagation, and however, the expression in (6) is valid under the assumptions described around (4), that is, correlated multipath propagation. The core DPD processing, described next, will thus focus on minimizing such beamformed distortion.

## B. DPD Principle and FD Processing

In order to be able to execute beam-level linearization, the DPD unit is placed prior to the BB precoder and thus acts in the FD. To this end, in order to linearize the far-field response of the array, the DPD must produce a predistorted signal with similar structure to that of the observable nonlinear distortion in (6), which is dependent on  $\Upsilon_{p,m}(n) = (d(n-m) \star F^{\text{eff}}(n-m)) |d(n-m) \star F^{\text{eff}}(n-m)|^{p-1}$ . Thus, placing the DPD directly to the subcarrier data symbol domain and learning the DPD to cancel the beamformed distortion of the form  $\alpha_{p,m}^{\text{tot}} \Upsilon_{p,m}(n)$  would create strong channel dependence on the DPD system through  $F^{\text{eff}}(\cdot)$ .

Consequently, in order to relax such channel dependency, we consider first splitting the BB precoder into two stages, motivated already by (4)—a common wideband stage corresponding to an impulse response  $F^{\text{eff}}(n)$  and an additional beamforming stage that applies the antenna-specific weights of the form  $e^{-j\beta_l}$ . The wideband precoder will be then placed prior to the DPD unit so that  $F^{\text{eff}}(n)$  is effectively considered in the generation of the DPD BFs, whereas the beamforming stage remains in the usual position before the OFDM waveform generation. It is also noted that in the case of pure LOS conditions, such channel dependency does not take place as  $F^{\text{eff}}(n) = \delta(n)$ .

Now, the wideband precoded sequence, expressed in FD, is given by

$$s[k] = \begin{cases} F^{\text{eff}}[k]d[k], & \text{for } -K_{\text{act}}/2 \leq k \leq K_{\text{act}}/2 - 1 \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

which can also include FD oversampling required for proper DPD operation. Assuming then an MP-type DPD processing, the actual DPD output signal, expressed also in the FD, reads

$$\tilde{s}[k] = s[k] + \sum_{p=1}^P \sum_{m=0}^M \gamma_{p,m}^* e^{-j2\pi mk/K} \Upsilon_{p,0}[k] \quad (8)$$

where  $\Upsilon_{p,0}[k]$  denotes the  $p$ th-order instantaneous BF at subcarrier  $k$ . It is noted that in the FD, the delayed BFs of the MP system are seen as per-subcarrier phase rotations of the form  $e^{-j2\pi mk/K}$  acting on the instantaneous BFs. Furthermore, the instantaneous BFs are calculated in the FD as

$$\begin{aligned}
\Upsilon_{1,0}[k] &= s[k] \\
\Upsilon_{2,0}[k] &= s[k] \star s^*[-k] \\
\Upsilon_{3,0}[k] &= s[k] \star s^*[-k] \star s[k] \\
\Upsilon_{p,0}[k] &= \Upsilon_{p-2,0}[k] \star \Upsilon_{2,0}[k]. \quad (9)
\end{aligned}$$

It is noted that since the DPD system in (8) builds on adding a low-power nonlinear cancellation signal in FD, we call it an injection-based DPD in the continuation.

Alternatively, as generating the FD instantaneous BFs through linear convolutions can be cumbersome, a more efficient approach is to calculate them through fast convolutions (FCs) instead. For instance,  $\Upsilon_{2,0}[k]$  can be calculated as

$$\Upsilon_{2,0}[k] = \text{DFT}^K \left( \text{IDFT}^K(s[k]) \cdot (\text{IDFT}^K(s[k]))^* \right) \quad (10)$$

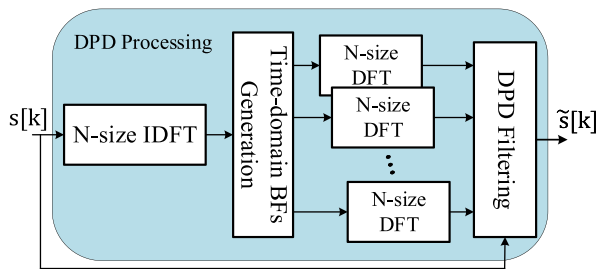


Fig. 3. Efficient implementation structure for FD DPD processing.

where  $K$  denotes the size of the transform. In the problem at hand, the sequence  $s[k]$  is convolved with itself, and hence, only one IDFT operation and as many DFTs as there are instantaneous nonlinear BFs are required. The overall DPD processing is graphically shown in Fig. 3, where the DPD filtering block executes the operation in (8) while also leveraging the FC approach for BF generation.

### C. Frequency Multiplexing and Out-of-Band Aspects

Executing FD DPD allows us to predistort the transmit signal at a subcarrier level. Consequently, different DPD filter orders can be applied based on specific FD masks. For instance, such mask can be defined based on scheduling and link adaptation decisions following the corresponding signal quality requirements applicable to different data modulation schemes. To this end, a filtered set of BFs is defined as follows:

$$\tilde{\Upsilon}_{p,m}[k] = \begin{cases} \Upsilon_{p,m}[k], & \text{if } k \in \mathcal{K}_p, \quad p = 1, 2, \dots, P \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

where  $\mathcal{K}_p$  defines the set of frequency bins, both inband and out-of-band (OOB), where the  $p$ th-order BFs are applicable. As for the OOB bins, the BW of the predistorting signal can be controlled very flexibly, e.g., so that it is the smallest one that allows fulfilling the OOB emissions mask. Concrete examples of such will be provided in Section VII along with the numerical results.

In general, the nonlinear distortion stemming from the PAs is beamformed in the direction defined by  $e^{-j\beta l}$ , and this applies not only to the passband distortion but also to the OOB emissions, as shown in Fig. 2. However, as the low-power DPD injection signal excites the PAs essentially linearly and the beamformer is classically only defined in the passband, the DPD signal is only naturally beamformed at the passband frequency bins. This results in the OOB distortion and the OOB DPD injection signal not experiencing the same effective beamforming responses. Consequently, in order for the DPD to be able to provide OOB linearization, the DPD injection signal outside the passband subcarriers must be beamformed too. To this end, the extended beamformer corresponding to the  $l$ th transmit chain is defined as

$$w_l[k] = e^{-j\beta l}, \quad \text{for } k \in \mathcal{K}_{\text{DPD}} \quad (12)$$

where  $\mathcal{K}_{\text{DPD}}$  denotes the set of frequency bins within which the DPD is executed. Finally, it is noted that in the case of

### Algorithm 1 FD DPD Learning

- 
- 1: **Inputs:**  $s[k]$ ,  $I$ ,  $\Upsilon$  and  $\mathbf{R}^{-1}$
  - 2: Set  $\boldsymbol{\gamma}(1) = \mathbf{0}$
  - 3: **for**  $i = 1$  to  $I$  **do**
  - 4:   Generate instantaneous BFs  $\Upsilon_{p,0}[k]$  through FC
  - 5:   Generate filtered BFs  $\tilde{\Upsilon}[k]$  based on a freq. mask
  - 6:   Generate memory BFs in the freq. domain
  - 7:   Inject the DPD signal as per (8)
  - 8:   Apply extended beamforming as per (12)
  - 9:   Execute waveform generation and transmission
  - 10:   Acquire  $z_{\text{fb}}[k]$  and calculate  $e[k]$  as per (14)
  - 11:    $\boldsymbol{\gamma}(i+1) \leftarrow \boldsymbol{\gamma}(i) - \mu \mathbf{R}^{-1} \Upsilon^T(i) \mathbf{e}^*(i)$
  - 12: **end for**
  - 13: **return**  $\boldsymbol{\gamma}(I)$
- 

two or more users being frequency-multiplexed, one needs to consider the DPD cancellation beams at the corresponding main-beam directions and frequency bins while otherwise following similar processing as that described in the single-user case.

### D. Combined Feedback-Based DPD Learning

In practice, the nonlinear characteristics of the individual PAs may change over time due to, for example, device aging or temperature drifts. Hence, a DPD tracking system is required. In order to acquire a proper signal for DPD learning that resembles the far-field nonlinear distortion, the so-called antibeaming and combining-based observation receiver is commonly considered [1], [3], [18], [30] and thus adopted also in this work. Considering that the antibeaming block is matched to the phase of the dominant tap of every antenna channel, its output is a linear combination of the PA output signals, which is expressed as

$$\begin{aligned} z_{\text{fb}}(n) &= \sum_{l=1}^L \sum_{p=1}^P \sum_{m=0}^M e^{j\beta l} y_l(n) \\ &= \sum_{p=1}^P \sum_{m=0}^M \alpha_{p,m}^{\text{tot}} \Upsilon_{p,m}(n). \end{aligned} \quad (13)$$

The feedback signal in (13) resembles the far-field beamformed signal at the receiver, shown in (6), except for the actual channel filtering that is common to both the DPD injection signal and the actual distortion stemming from the TX, and does not influence the parameter learning [18].

The output of the feedback receiver is then utilized to obtain the FD error signal for the actual parameter learning. This FD error signal can be expressed as

$$e[k] = z_{\text{fb}}[k] - g_{\text{tot}} s[k], \quad \text{for } k \in \mathcal{K}_{\text{DPD}} \quad (14)$$

where  $g_{\text{tot}}$  is the effective complex linear gain of the array and  $e[k]$  contains information about the residual distortion at the subcarrier level. In addition, the information in  $e[k]$  can be exploited to track the power of the residual distortion across the frequency bins to ensure that a given frequency mask allows fulfilling the signal quality requirements and tuning

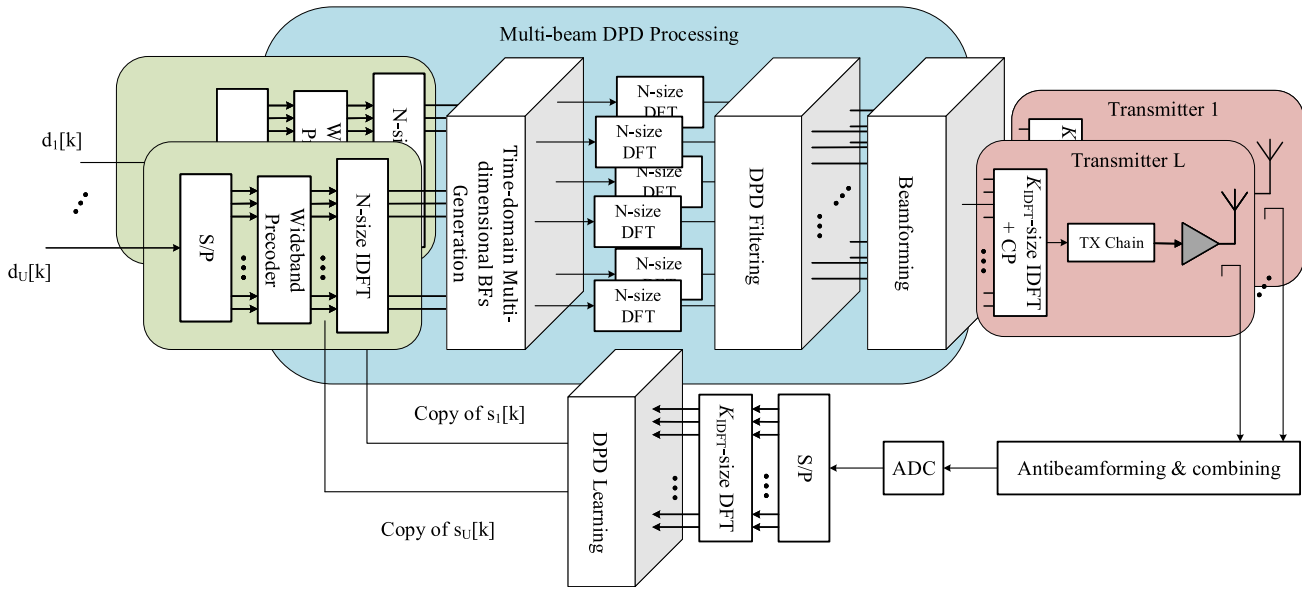


Fig. 4. Fully digital MIMO TX architecture with beam-level DPD and multiuser transmission.

the DPD parameters accordingly. The corresponding self-orthogonalized least-mean-squares (LMS) learning rule [31] for the DPD weights  $\boldsymbol{\gamma}$ , expressed here at block iteration  $i$ , reads then

$$\boldsymbol{\gamma}(i+1) = \boldsymbol{\gamma}(i) - \mu \mathbf{R}^{-1} \boldsymbol{\Upsilon}^T(i) \mathbf{e}^*(i) \quad (15)$$

where the vector  $\mathbf{e}$  contains the error signal samples  $e[k]$  of the subcarriers spanned by the DPD and  $\mathbf{R}^{-1}$  is the inverse of the sample estimate of the correlation matrix calculated as  $\mathbf{R} = \boldsymbol{\Upsilon}^H \boldsymbol{\Upsilon}$ .

The FD DPD learning algorithm is summarized in Algorithm 1, where  $I$  denotes the number of LMS block iterations.

#### IV. PROPOSED MULTIUSER BEAM-LEVEL DPD

In this section, the extension of the proposed DPD structure to the more general multiuser case with  $U > 1$  is provided. The corresponding TX architecture is shown in Fig. 4.

##### A. Spatial Characteristics of Nonlinear Distortion Under Multiuser Transmission

Without loss of generality, and in order to emphasize the mathematical tractability of the expressions, we consider a two-user scenario, i.e.,  $U = 2$  and, thus,  $u = 1, 2$ . Considering now first that the DPD processing unit is OFF, every data stream  $d_u(n)$  will be shaped by its corresponding wideband precoders  $F_u(n)$ . Then, the precoded streams will be beamformed by the beamforming unit of dimensions  $L \times U$  with entries  $e^{-j\beta_{l,u}}$ . Such beamforming results in a linear combination of the precoded data streams, forming the  $l$ th PA input signal of the form

$$\begin{aligned} x_l(n) &= d_1(n) \star F_1(n) e^{-j\beta_{l,1}} + d_2(n) \star F_2(n) e^{-j\beta_{l,2}} \\ &= \Upsilon_{1,1,0}(n) e^{-j\beta_{l,1}} + \Upsilon_{2,1,0}(n) e^{-j\beta_{l,2}} \end{aligned} \quad (16)$$

where the first subscript in  $\Upsilon_{u,1,0}(n)$  refers to the user index. The corresponding  $l$ th PA output signal then reads

$$\begin{aligned} y_l(n) &= \sum_{p=1}^P \sum_{m=0}^M \alpha_{p,m}^l \left( \Upsilon_{1,1,m}(n) e^{-j\beta_{l,1}} + \Upsilon_{2,1,m}(n) e^{-j\beta_{l,2}} \right) \\ &\quad \times \left| \Upsilon_{1,1,m}(n) e^{-j\beta_{l,1}} + \Upsilon_{2,1,m}(n) e^{-j\beta_{l,2}} \right|^{p-1} \\ &= \sum_{p=0}^P \sum_{k=0}^{p+1} \sum_{v=0}^p \sum_{m=0}^M \alpha_{p,m}^l \Upsilon_{2,1,m}^k(n) \Upsilon_{2,1,m}^{*v}(n) e^{-j(k-v)\beta_{l,2}} \\ &\quad \times \Upsilon_{1,1,m}^{p+1-k}(n) \Upsilon_{1,1,m}^{*(p-v)}(n) e^{-j(v-k+1)\beta_{l,1}}. \end{aligned} \quad (17)$$

It is noted that similar nonlinear models have been utilized to characterize digital TXs affected by nonlinear crosstalk [8], [24], [25], to model concurrent dual-band TXs [32], to model the passive-intermodulation present in frequency-division duplexing radios [33], or to model the different intermodulation products in fully connected hybrid beamforming TX architectures [13]. In addition, we note that the model in (17) can be generalized to  $U > 2$  users by considering a multivariable polynomial model of the form of [13, eq. (12)] and by further considering the complex exponentials defining the beam directions into the model.

As can be seen through (17), IMD between the different signal streams appears, which also results in new beamforming directions. For simplicity, we focus next on the 3rd-order distortion terms, which are listed in the following:

$$\begin{aligned} &\Upsilon_{1,1,m}(n) \left| \Upsilon_{1,1,m}(n) \right|^2 e^{-j\beta_{l,1}} \\ &\Upsilon_{2,1,m}(n) \left| \Upsilon_{2,1,m}(n) \right|^2 e^{-j\beta_{l,2}} \\ &\Upsilon_{1,1,m}(n) \left| \Upsilon_{2,1,m}(n) \right|^2 e^{-j\beta_{l,1}} \\ &\Upsilon_{2,1,m}(n) \left| \Upsilon_{1,1,m}(n) \right|^2 e^{-j\beta_{l,2}} \\ &\Upsilon_{1,1,m}(n) \Upsilon_{1,1,m}(n) \Upsilon_{2,1,m}^*(n) e^{-j(2\beta_{l,1} - \beta_{l,2})} \\ &\Upsilon_{2,1,m}(n) \Upsilon_{2,1,m}(n) \Upsilon_{1,1,m}^*(n) e^{-j(2\beta_{l,2} - \beta_{l,1})}. \end{aligned} \quad (18)$$

The distortion terms above can be further grouped based on the direction the distortion is beamformed to. For the two-user case and 3rd-order nonlinearities, there are four effective beamforming directions:

- 1)  $e^{-j\beta_{1,l}}$ , which corresponds to the direction of user one;
- 2)  $e^{-j\beta_{2,l}}$ , which corresponds to the direction of user two;
- 3)  $e^{-j(2\beta_{1,l}-\beta_{2,l})}$  and  $e^{-j(2\beta_{2,l}-\beta_{1,l})}$ , which correspond to the directions defined by the intermodulation between the two signal streams.

As a concrete example, considering that the two users are served with equal power and are located at example directions of  $12^\circ$  and  $-15^\circ$ , the far-field beampatterns illustrating the above indicated beamforming directions, in addition to higher order ones, for a 100-antenna TX are shown in Fig. 5(a). The resulting 3rd-order intermodulation directions are at  $39^\circ$  and  $-42^\circ$  and dominate over the higher order ones, becoming a nonnegligible source of distortion. In black, the total radiated signal is illustrated, while the blue, red, and gray beampatterns correspond to the total linear and nonlinear distortion, passband plus adjacent channel nonlinear distortion, and adjacent channel distortion, respectively. It is important to note that the two intended users, located at  $12^\circ$  and  $-15^\circ$ , are served at the same time/frequency resources in the allocated channel as they are being spatially multiplexed. This implies that the distortion that they observe corresponds to the blue beampattern. On the other hand, potential victim receivers located at the intermodulation directions,  $39^\circ$  and  $-42^\circ$ , are operating at the adjacent channel; consequently, only the distortion emissions corresponding to the gray beampattern concern them. The extension of the previous DPD of Section III to this type of multiuser cases will build on introducing the necessary processing such that all the dominant beams can be linearized at the applicable frequency bins.

Furthermore, in real networks, it is unlikely that all users are served with similar transmit powers as their distance to the BS can vary considerably. For instance, there could be one user at the cell edge that would require higher TX power, while the rest could be close to the BS, thus necessitating lower power for good link quality. In this case, the spatial characteristics of the nonlinear distortion resemble those of the single-user case, meaning that there is only one dominant direction where OOB emissions are strongly beamformed to, which corresponds to that of the dominant user [15]. In Fig. 5(b), an example case where six users are spatially multiplexed and two of them are transmitted with 18 dB more power than the other four is considered. These high-power users are referred to as the dominant users in the following. As it can be observed, the beampattern resembles that of the two-user case shown in Fig. 5(a), implying that there are four dominant directions where OOB distortions are beamformed to. Hence, the BS can be serving multiple users while only requiring few directions to be linearized. Thus, while the complexity of the proposed DPD system is user dependent, the characteristics of the distortion can be exploited to reduce the amount of beams to be linearized, particularly those at the intermodulation directions that grow rapidly with the number of spatially multiplexed users. It is, however, noted that despite the number of relevant intermodulation directions that depend

on the number of dominant users, it may still be necessary to provide linearization in the direction of the low-power users to enhance their inband signal quality. This will be further discussed along the numerical results in Section VII.

Fig. 6 shows further the case in which six users are all served with equal/similar power. Naturally, the amount of spatial directions toward which the nonlinear distortion is beamformed increases due to the richer set of intermodulation products. Interestingly, the nonlinear distortion becomes increasingly isotropic, and this effect becomes more apparent as the number of user increases. More importantly, the relative level of the nonlinear distortion—the red and the gray beampatterns—with respect to the passband power decreases. For further details, the interested reader is kindly referred to [15], where the authors proved that as the number of spatially multiplexed users grows, the radiated nonlinear distortion becomes increasingly omnidirectional and lower. Hence, in the fairly unlikely event that the base station is spatially multiplexing really a large number of users with the same relative powers, the nonlinear distortion could be sufficiently low so that running a DPD might be completely avoided. In such a case, the DPD could be momentarily switched OFF until it is again required.

### B. Multiuser DPD Structure

In the case of multiuser transmission, the DPD has to consider BFs with the same structure as the nonlinear terms in (17). As a result, the DPD unit becomes multidimensional with dimensions  $B \times U$ , where  $B$  is the total number of dominant beams to be linearized. For instance, assuming that the list of relevant distortion directions is those given in (18), the DPD would have two inputs and four outputs. In general, the most relevant distortion beamforming directions are the intended users' ones, in which sufficiently good signal quality must be ensured. On the other hand, at the intermodulation directions, potential victim users that might be located there would be operating at the adjacent channels, and hence, only OOB linearization across a reduced BW may be needed to fulfill the FR2 adjacent channel leakage ratio (ACLRL) requirements. The FD processing offers natural tools for tailoring the DPD frequency contents accordingly.

In order to provide linearization toward the intended users, the DPD needs to inject to the different signal streams the relevant distortion terms that are beamformed toward such directions. Continuing with the previous two-user scenario, these terms can be expressed as

$$\begin{aligned} \tilde{s}_1(n) &= s_1(n) + \sum_{p=1}^P \sum_{q=1}^{P-p+1} \sum_{m=0}^M \gamma_{p,q,m}^1 \Upsilon_{1,1,m}(n) \\ &\quad \times |\Upsilon_{1,1,m}(n)|^{2(p-1)} |\Upsilon_{2,1,m}(n)|^{2(q-1)} \quad (19) \\ \tilde{s}_2(n) &= s_2(n) + \sum_{p=1}^P \sum_{q=1}^{P-p+1} \sum_{m=0}^M \gamma_{p,q,m}^2 \Upsilon_{2,1,m}(n) \\ &\quad \times |\Upsilon_{2,1,m}(n)|^{2(p-1)} |\Upsilon_{1,1,m}(n)|^{2(q-1)} \quad (20) \end{aligned}$$

which are of the form of  $2 \times 2$  parallel Hammerstein models [8], [24]. Then, each of the signals  $\tilde{s}_1(n)$  and  $\tilde{s}_2(n)$  will

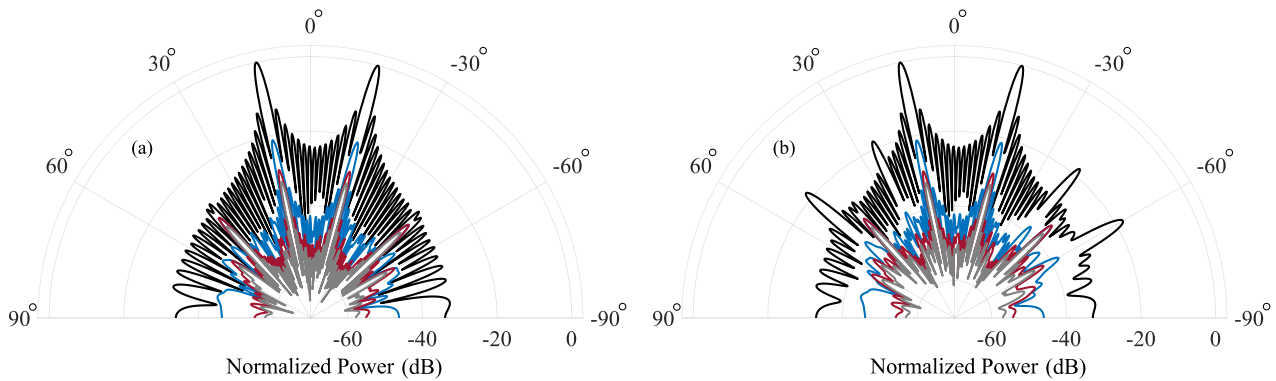


Fig. 5. Example far-field beam patterns for total radiated signal (black), total distortion (blue), passband nonlinear distortion (red), and OOB distortion (gray) for (a) two spatially multiplexed users served with equal power and (b) six spatially multiplexed users out of which two are dominant users; 100 antennas and PAs are considered.

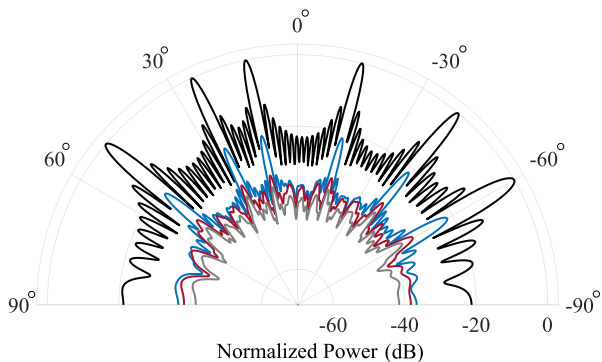


Fig. 6. Far-field beam patterns for total radiated signal (black), total distortion (blue), passband nonlinear distortion (red), and OOB distortion (gray) for six users served with equal power; 100 antennas and PAs are considered.

be beamformed toward the directions defined by  $e^{j\beta_{1,l}}$  and  $e^{j\beta_{2,l}}$ , respectively, across the applicable set of frequency bins, similar to the single-user case.

In order to provide linearization toward the intermodulation directions, the DPD has to again inject a proper set of corresponding distortion terms. Bearing in mind that at the intermodulation directions, there is no data stream to transmit, and that potential victim receivers located at such locations are operating at the adjacent channel(s), only nonlinear BFs are required. Considering for simplicity that we aim at canceling the distortion beamformed toward  $e^{j(2\beta_{1,l}-\beta_{2,l})}$  and  $e^{j(2\beta_{1,l}-\beta_{2,l})}$ , the corresponding relevant DPD terms are of the form

$$\begin{aligned} \tilde{s}_{(2\beta_1-\beta_2)}(n) &= \sum_{m=0}^M \Upsilon_{1,1,m}^2(n) \Upsilon_{2,1,m}^*(n) \\ &+ \Upsilon_{1,1,m}^2(n) \Upsilon_{2,1,m}^*(n) |\Upsilon_{1,1,m}(n)|^2 \\ &+ \Upsilon_{1,1,m}^2(n) \Upsilon_{2,1,m}^*(n) |\Upsilon_{2,1,m}(n)|^2 + \dots \end{aligned} \quad (21)$$

$$\begin{aligned} \tilde{s}_{(2\beta_2-\beta_1)}(n) &= \sum_{m=0}^M \Upsilon_{2,1,m}^2(n) \Upsilon_{1,1,m}^*(n) \\ &+ \Upsilon_{2,1,m}^2(n) \Upsilon_{1,1,m}^*(n) |\Upsilon_{1,1,m}(n)|^2 \\ &+ \Upsilon_{2,1,m}^2(n) \Upsilon_{1,1,m}^*(n) |\Upsilon_{2,1,m}(n)|^2 + \dots \end{aligned} \quad (22)$$

while the corresponding higher order terms can be generated from the general expression in (17).

In addition to generating the right injection terms, the intermodulation cancellation signals need to be beamformed toward their respective directions of interest. Hence, the beamforming block is extended from  $L \times U$  dimensions to  $L \times B$ , i.e., it contains the weights required to beamform the predistorted signals toward the intended users and toward the considered intermodulation directions whose distortion is to be compensated for. In this case, the corresponding input signal to the  $l$ th PA reads

$$\begin{aligned} \tilde{x}_l(n) &= \tilde{s}_1(n) e^{-j\beta_{1,l}} + \tilde{s}_{(2\beta_1-\beta_2)}(n) e^{-j(2\beta_{1,l}-\beta_{2,l})} \\ &+ \tilde{s}_2(n) e^{-j\beta_{2,l}} + \tilde{s}_{(2\beta_2-\beta_1)}(n) e^{-j(2\beta_{2,l}-\beta_{1,l})}. \end{aligned} \quad (23)$$

In the considered example case, one can think of such multistream DPD processing as four independent multidimensional DPD units, where each of the units is responsible for linearizing a specific beam. Hence, the learning approach will be similar to that of the single-user case with some minor distinctions, which are detailed in the following. For clarity, it is noted that the parallel DPD structure shown in Fig. 1(a) does not require such multidimensional BFs. This is because in such architecture, the DPD takes as input the spatially multiplexed signal  $x_l(n) = d_1(n) \star F_{l,1}(n) + d_2(n) \star F_{l,2}(n)$ , and hence, the DPD will implicitly generate such intermodulation products through more ordinary SISO BFs.

Finally, it is noted that the DPD IMD cancellation terms in (23) do not result in any essential loss of linear beamforming gain since such beams are of very low power and are generated independently of the linear signals. Furthermore, the IMD cancellation beams only consider nonlinear BFs at the frequency bins corresponding to the adjacent channel frequencies.

### C. Multiuser DPD Learning

In order to train the multistream DPD, the observation receiver should provide proper reference signals that resemble the far-field nonlinear distortion at the directions of interest. One natural way of doing so is by time multiplexing the different beam observations, as the individual per-stream DPDs can be trained independently. To that end, the phases of the antibeamforming block will be changed sequentially so that



TABLE I

DPD MAIN PATH PROCESSING COMPLEXITY PER OFDM SYMBOL AND DPD LEARNING COMPLEXITY PER BLOCK ITERATION FOR THE PROPOSED BEAM-BASED DPD AND FOR MORE ORDINARY TIME-DOMAIN DPD OPERATING ON PER-TX BASIS

		Proposed Beam-level FD DPD [FLOPs]	Per-TX TD DPD [FLOPs]
<b>DPD</b>	IFFT	$U(K_{\text{act}}(\log_2(K_{\text{act}}) - 3) + 3K_{\text{act}}(\log_2(K_{\text{act}}) - 1) + 8)$	–
	FFT	$(N_{\text{IBF}} - U)(K_{\text{DPD}}(\log_2(K_{\text{DPD}}) - 3) + 3K_{\text{DPD}}(\log_2(K_{\text{DPD}}) - 1) + 8)$	–
	BF gen.	$(2N_{\text{IBF}} + 2 - U)K_{\text{DPD}} + 4M \sum_{p=1}^P  \mathcal{K}_p $	$L(2N_{\text{IBF}} + 1)N_{\text{OFDM}}$
	Filtering	$8(M + 1) \sum_{p=1}^P  \mathcal{K}_p  - 2K_{\text{DPD}}$	$L(8N_{\text{BF}} - 2)N_{\text{OFDM}}$
<b>Learning</b>	FFT	$B(K_{\text{DPD}}(\log_2(K_{\text{DPD}}) - 3) + 3K_{\text{DPD}}(\log_2(K_{\text{DPD}}) - 1) + 8)$	–
	DPD est.	$BN_{\text{CL}}N_{\text{BF}}(4N_{\text{BF}} + 6)$	$N_{\text{CL}}N_{\text{BF}}(4N_{\text{BF}} + 6)L$

enough samples are recorded for each direction. Then, each of the stream DPDs can be trained following the same principle as in the single-user case described in Algorithm 1. In addition, the error signal for the intermodulation directions is generated as

$$e_{\text{int}}[k] = \tilde{z}_{\text{fb}}[k] \quad (24)$$

i.e., directly as the output of the observation receiver. This is because, at the intermodulation directions, all the radiated energy is distortion.

## V. IMPACT OF ARRAY CROSSTALK

Crosstalk is an additional hardware impairment commonly encountered in array TXs and is known to spoil the performance of classical SISO DPD systems [30]. Crosstalk in fully digital TXs has been extensively studied in current art, and various multidimensional DPD structures have been proposed to cope with it [8], [9], [24], [25]. As the signal models in the presence of crosstalk are well established, in this section, a brief discussion on the implications of crosstalk in the beam-level and parallel DPD architectures is provided, while the effects of the crosstalk are then further analyzed along the numerical results.

Considering first that every PA is affected by PA input crosstalk [8], [24], the input signal to the  $l$ th PA reads

$$x_l^{\text{ct}}(n) = x_l(n) + \sum_{\substack{i=1 \\ i \neq l}}^l c_{i,l} x_i(n) \quad (25)$$

where  $c_{i,l}$  is the complex PA input coupling coefficient between the  $i$ th and  $l$ th TX chains. It is noted that the PA input crosstalk may result in new intermodulation products different to those in (16), as the signals  $x_i(n)$  are beamformed according to  $e^{-j\beta_{u,i}}$ . However, the relative level of the coupled signal with respect to the transmit signal at the corresponding TX path is in general between  $-10$  and  $-20$  dB [8], [9], [24], [25], and hence, they will not result in dominant intermodulation directions, similar to the case in which some users are served with significantly higher power than the others.

Similarly, the PA output crosstalk is modeled as [8], [24]

$$y_l^{\text{ct}}(n) = y_l(n) + \sum_{\substack{i=1 \\ i \neq l}}^l b_{i,l} y_i(n) \quad (26)$$

where  $b_{i,l}$  is the complex PA output coupling coefficient between the  $i$ th and  $l$ th TX chains.

Despite the crosstalk being a physical phenomenon that originates independently of the adopted DPD structure, there are very important differences depending on which architecture is adopted. In the classical parallel DPD configuration, the DPD provides linearization at the PA level, and hence, any unmodeled distortion present at the PA output is captured by the observation receiver hindering the parameter learning and thereon degrading the performance [8], [9], [18], [24], [25]. Hence, a multidimensional DPD structure per PA is required in such a case so that also the crosstalk terms can be accounted for. To this end, in a highly simplistic example case of two antenna units, the BFs have the same structure as those reported in Section IV. On the other hand, the proposed beam-level DPD tackles the problem from a beamformed channel perspective, implying that the output of the observation receiver provides a coherent sum of the antenna signals, which is then provided to the DPD learning engine. However, the resulting crosstalk terms are summed noncoherently, and thus, they are diluted in comparison to the distortion in (13). Hence, such combined feedback configuration provides robustness against unmodeled nonbeamformed distortion [18]. This is demonstrated through the extensive numerical results in Section VII. In addition, in the proposed beam-level DPD structure, the BFs do not build on any specific beamforming direction opposed to the parallel DPD structure, in which the BFs are built in terms of the beamformed streams  $x_l(n)$ . Hence, if any of the crosstalk terms get beamformed toward any of the considered dominant beam directions, the proposed beam-level DPD would be capable of suppressing them. This is because the beam-level DPD already considers the intermodulation BFs that are of the form shown in (19)–(22), which in turn are similar to those commonly utilized to compensate for crosstalk in classical DPD configurations [8], [9], [24], [25].

## VI. COMPLEXITY ANALYSIS

In this section, the complexity entailed by the proposed beam-level DPD is analyzed and benchmarked against that of the parallel DPD reference configuration in terms of floating-point operations (FLOPs). Both DPD architectures are assumed to utilize a closed-loop (CL) parameter learning architecture based on the learning rule in (15) executed either

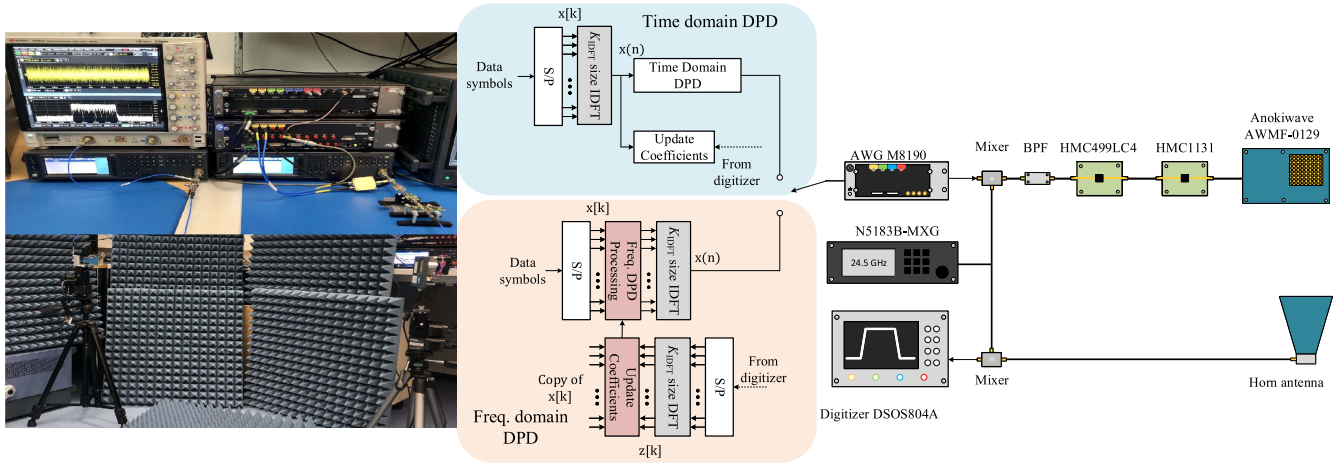


Fig. 7. mmWave measurement setup and BB processing for the proposed beam-level DPD and the reference TD DPD solution considering a phased-array TX.

in the frequency or in the time domain. For convenience, the notations utilized in this section are summarized next. To this end,  $K_{\text{act}}$  and  $K_{\text{DPD}}$  denote the number of frequency bins in the passband and within the frequency range spanned by the DPD, respectively. In addition,  $N_{\text{IBF}}$  and  $N_{\text{BF}}$  refer to the numbers of the instantaneous BFs and the total BFs, respectively, while  $|\mathcal{K}_p|$  denotes the cardinality of the set  $\mathcal{K}_p$ . Finally,  $N_{\text{CL}}$  denotes the samples per block iteration,  $U$  is the number of users,  $B$  is the number of dominant beams, and  $L$  is the total number of the RF chains.

The basic assumptions adopted in the complexity analysis are detailed in the following.

- 1) One complex multiplication is assumed to cost six FLOPs, while one complex-real multiplication and one complex sum both cost two FLOPs.
- 2) The polynomial BFs are built utilizing the so-called Horner's rule [34].
- 3) An individual  $K_{\text{FFT}}$ -point FFT/IFFT transform requires  $K_{\text{FFT}} \log_2 K_{\text{FFT}} - 3K_{\text{FFT}} + 4$  real multiplications and  $3K_{\text{FFT}} \log_2 K_{\text{FFT}} - 3K_{\text{FFT}} + 4$  real additions [35].
- 4) CL DPD structures utilize the same BFs in both the DPD main path and in the parameter learning. Consequently, the learning can reuse the BFs already available from the main path transmission whenever needed.
- 5) In the proposed solution, the instantaneous BFs are generated in the time domain, and in order to do so, one IFFT per signal stream and as many FFTs as instantaneous nonlinear BFs, there are need to be computed. The IFFTs are computed on  $s[k]$ , which only has  $K_{\text{act}}$  nonzero samples, while the FFTs operate on the oversampled sequences. On the other hand, the memory BFs are calculated directly in the FD as indicated in (8).
- 6) In the per-TX reference solution, the BFs are generated directly in the time domain without involving pairs of DFT transforms. In addition, generating the memory BFs in the time domain does not involve any FLOPs, as they are delayed versions of the instantaneous BFs.
- 7) As the reference solution operates in the time domain, every DPD unit needs to filter all the  $N_{\text{OFDM}}$  samples within an OFDM symbol. On the other hand, the proposed DPD, which operates in the FD, has to filter

the subcarrier symbols that lie within the corresponding range of operation of the  $p$ th order BFs. For instance, the linear BFs are only applied to the passband subcarriers.

The obtained generic complexity expressions are gathered in Table I, while exact complexity numbers are provided for specific evaluation scenarios in Section VII. In Table I, BF Gen. refers to the complexity of generating the required BFs, the filtering complexity corresponds to computing (8), and DPD est. refers to the complexity stemming from executing one block iteration of the learning rule in (15). In the parallel per-TX DPD configuration, the learning is executed  $L$  times, while in the proposed beam-level DPD solution, it is executed  $B$  times.

## VII. NUMERICAL RESULTS

In this section, extensive numerical results in terms of both actual RF measurements as well as simulation-based experiments are reported and analyzed in order to demonstrate the capabilities of the proposed beam-level DPD solutions, together with the flexibility achieved through operating in the FD. Exact complexity numbers are reported for every conducted experiment considering the analysis in Section VI. The proposed solutions are benchmarked against the classical parallel DPD configuration operating in the time domain, and both architectures are considered to utilize a CL architecture with the same self-orthogonalized learning principle in (15). The test waveform considered in the experiments is a 5G NR standard-compliant 200-MHz OFDM waveform composed of  $K_{\text{act}} = 3168$  active subcarriers and a sampling rate of 1.2288 Gsamples/s. The spectral containment of the digital TX waveform is enhanced, compared to plain CP-OFDM, using the well-known weighted-overlap-and-add (WOLA) type of windowing in the time domain.

### A. RF Measurement Results: Single-User Beam-Level DPD and FD DPD Principle

The mmWave experimental testbed is shown in Fig. 7 and is composed of the following elements: a Keysight M8195A arbitrary waveform generator that is utilized to generate the TX IF signal centered at 3.5 GHz. A Keysight N5183B-MXG

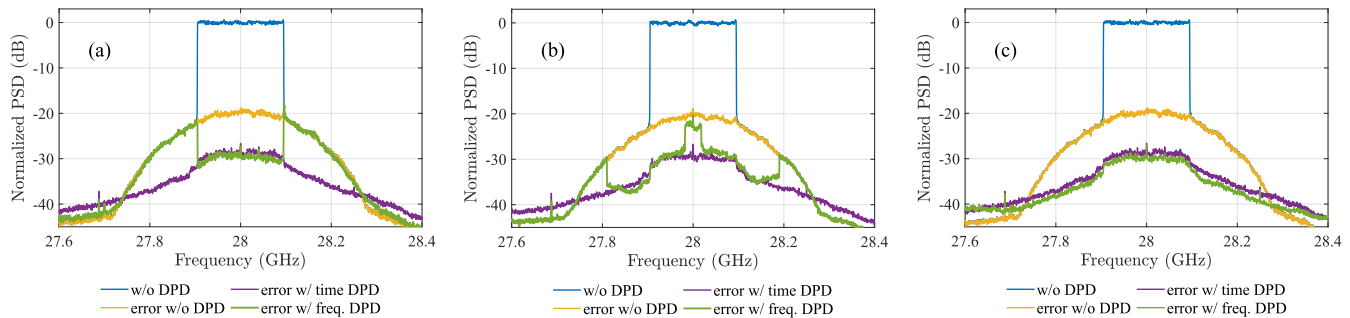


Fig. 8. Measured OTA combined spectra w/o DPD and with the proposed FD DPD with different frequency masks: (a) passband DPD, (b) frequency-selective passband + partial OOB mask, and (c) full-band DPD. TD DPD performance is also shown for reference. EIRP  $\approx +43$  dBm.

signal generator running at 24.5 GHz is used to generate the local oscillator signal that, together with a mixer (Marki Microwave T3-1040), is utilized for upconverting the IF signal to the desired carrier frequency of 28 GHz at TX. Then, the modulated RF waveform at 28 GHz is amplified with one HMC499LC4 and one HMC1131 driver amplifier that allow us to feed the AWMF-0129 phased array with sufficient input power. The transmit signal then propagates OTA and is captured by a horn antenna. After downconversion to IF, the IF signal is digitized and taken to BB by means of a DSOS804A oscilloscope. Finally, the samples are processed in a host PC running MATLAB, where the different DPD models are trained.

AWMF-0129 is a 64-element phased-array executing digitally controlled analog beamforming. It is representative of the single-user case detailed in Section III, with spatial emissions corresponding to those shown in Fig. 2. The following experiment serves as a basic proof of concept of the proposed beam-level DPD and its operation principle in the FD with state-of-the-art HW. The FD DPD follows the same principle as that described in Section III, i.e., the output of the observation RX is taken to FD and is then compared against the sequence  $s[k]$ —in this rank-1 single-user case unprecoded—to generate the error signal. Then, the DPD is executed prior to the waveform generation stage. As the considered array HW does not support the antibeaming and combining module to collect the observations for parameter estimation, the actual OTA received signal is directly used to train the DPD filter. It is noted that the considered phased array has a single RF input, which is then split accordingly to execute analog beamforming. Consequently, there are no capabilities to conduct digital beamforming, and a single DPD unit is responsible for linearizing the far-field beamformed distortion [1], [2] for both TD and FD implementations, i.e., they operate at beam level. The overall measurement setup and BB processing are shown in Fig. 7.

The BFs utilized in this experiment correspond to a classical MP structure, and both DPD solutions employ  $P = 7$  and  $M = 5$ , yielding a total of 24 BFs when considering only the odd-order nonlinearities. The size of the  $N$ -size transforms shown in Fig. 2 is 8192 for the two first linearization scenarios and 20480 for the last one, while the size of the final IDFT is varied based on the BW of the predistorter; specifically,  $K_{\text{IDFT}} = 4096, 8192, \text{ or } 20480$ . The measurement results in terms of the

TABLE II  
MEASURED OTA ACLR AND EVM FOR THE MMWAVE  
PHASED-ARRAY EXPERIMENTS IN FIG. 8

DPD Scenario	TRP ACLR (dBc)	EVM (%)
w/o DPD	26.61	10.04
freq. DPD, Fig. 9(a)	26.61	3.18
freq. DPD, Fig. 9(b)	30.22	3.37 / 7.91
freq. DPD, Fig. 9(c)	32.34	3.18

measured spectra are gathered in Fig. 8, showcasing several alternative scenarios of interest, with their corresponding total radiated power (TRP)-based ACLR and error vector magnitude (EVM) values being gathered in Table II.

In scenario [Fig. 8(a)], only passband linearization is pursued, providing sufficient signal quality to support up to 256-QAM modulation across the whole channel BW. In this case, the DPD filter is only active for  $k \in \{-1584, \dots, 1583\}$ . However, as the TRP-based ACLR is not fulfilled, some OOB linearization is required. To that end, a different frequency mask is adopted in Fig. 8(b), where the DPD filter is now active for  $k \in \{-3168, \dots, 3167\}$ . As can be observed, the FD DPD is now capable of fulfilling the OOB emission mask while being able to reduce the BW of the predistorted signal substantially compared to the ordinary TD DPD approach. In addition, this example also considers a frequency-selective mask in the passband, in which a third-order DPD filter is applied at the frequency bins  $k \in \{-300, \dots, 299\}$  and seventh order elsewhere. This example could correspond to a real scenario in which one user is served at the center of the channel utilizing 64-QAM modulation, while other users are served at the edges of the channel with 256-QAM modulation. The FD DPD is thus capable of conveniently adapting its linearization performance based on instantaneous user quality demands. Finally, the final evaluation scenario shown in Fig. 8(c) considers full-band linearization for the FD DPD, being able to provide similar linearization performance to its TD counterpart.

The exact complexity numbers corresponding to full-band and passband linearization scenarios are gathered in Table III. Such values are calculated for  $U = 1$ ,  $B = 1$ ,  $L = 1$ , and  $N_{\text{OFDM}} = 22528$ . As it can be observed, despite the TD DPD reference solution consisting of only a single DPD unit in this

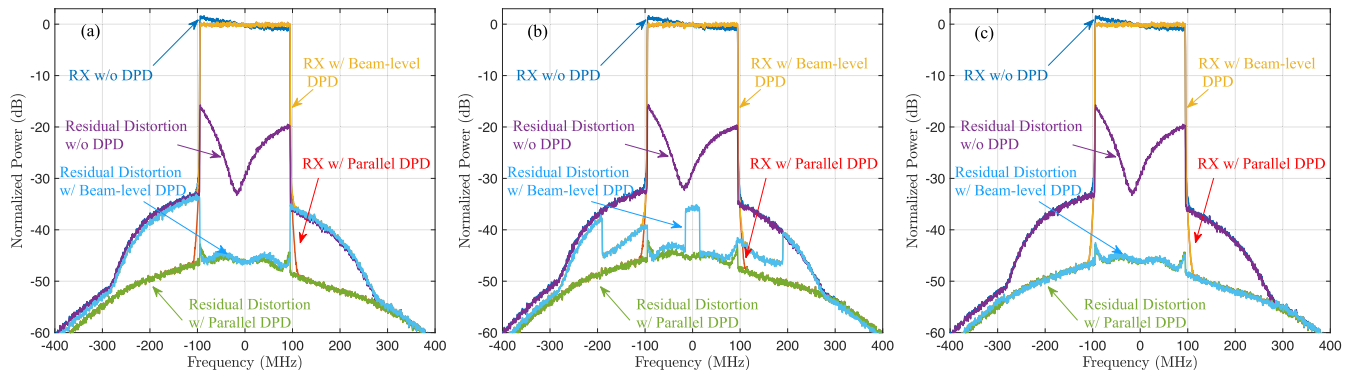


Fig. 9. RX spectra at the intended RX w/o DPD and with the parallel DPD approach and with the proposed FD DPD with different frequency masks. (a) Passband DPD. (b) Frequency-selective mask. (c) Full-band DPD. This simulation case corresponds to the single-user digital MIMO transmission scenario.

TABLE III

DPD MAIN PATH PROCESSING COMPLEXITY PER OFDM SYMBOL AND DPD LEARNING COMPLEXITY PER BLOCK ITERATION FOR THE SINGLE-USER CASE AND FULL-BAND/PASSBAND LINEARIZATION FOR THE MMWAVE PHASED-ARRAY EXPERIMENT. THE COMPLEXITY OF THE PER-TX TIME-DOMAIN DPD IS SHOWN FOR REFERENCE

		Beam-level FD DPD [MFLOPs]	Per-TX TD DPD [MFLOPs]
DPD	IFFT	0.1283	—
	FFT	3.1511 / 1.1305	—
	BF gen.	1.7349 / 0.3779	0.2028
	Filtering	3.6180 / 0.7096	5.0012
	<b>Total</b>	8.6325 / 2.3464	5.2040
Learning	FFT	1.0504 / 0.1284	—
	DPD est.	66.0800	66.0800
	<b>Total</b>	67.1304 / 66.2084	66.0800

phased-array scenario, the proposed passband FD DPD can still offer a significant reduction of computational complexity. The main reason for this is that the FD DPD is applied to the subcarrier symbols, while the TD DPD predistorts every high-rate transmit sample. This alone is already a remarkable implementation benefit. On the other hand, the complexity of the full-band DPD is larger than that of the time-domain DPD due to the complexity associated with the large-size IFFT/FFT transforms. Nevertheless, the complexity reduction will be significantly larger in the context of fully digital MIMO TXs, even in the full-band DPD case when implementing the proposed beam-based DPD solution, as it is shown in the following experiments.

### B. Simulation Results With Digital MIMO TX

Due to the lack of access to a digital large array MIMO TX, the capabilities and operating principle of the proposed beam-level DPD for the single-user and multiuser cases are showcased next through comprehensive simulation results. The evaluation scenario builds on the clustered mmWave channel model described in Section II, containing  $C = 5$  clusters each with  $R = 3$  rays. It is assumed that there is always a direct path between the TX and the users, with a Rician  $K$  factor of 10 dB. The maximum considered excess delay is

60 ns, which is well in-line with the assumptions in [36]. On the other hand, in order to assess the spatial-domain emissions, different far-field beam patterns, similar to those already illustrated along this paper, are utilized. In order to draw such far-field beam patterns, ideal LOS propagation conditions are assumed, while specific snapshot spectra considering the more realistic wideband propagation scenario with multipath components involved are also shown. We consider a fully digital MIMO TX with  $L = 100$  antenna units and PAs, which serves a varying amount of users so that the different scenarios discussed in Section IV can be addressed. The PA models considered in these simulations correspond to samples of HMC943APM5E ICs measured at 28 GHz under the stimulus of the OFDM waveform detailed at the beginning of this section. The measured PA models are available as supplementary files in [18]. The test waveform is the same as the one considered in the previous RF experiments.

1) *Single-User Case:* When the base station serves as a single user, the spatial characteristics of the nonlinear distortion correspond to those shown in Fig. 2, i.e., there is a single dominant direction/beam like in the previous RF experiment. This is the most simple scenario that can be addressed by the proposed beam-level DPD and makes it possible to obtain the largest complexity reduction, as only a single beam needs to be linearized—opposed to utilizing the 100 DPD units in the reference parallel DPD configuration. Naturally, a similar reduction in the learning complexity can also be obtained as only a single DPD is to be trained rather than 100. Both the proposed beam-based DPD and the parallel DPD configuration are considered to employ up to 5th order BFs (odd-order nonlinearities only) and  $M = 2$ , that is, a main tap and two delay taps. This corresponds to  $N_{\text{IBF}} = 3$  and  $N_{\text{BF}} = 9$ . The linearization performance in terms of spectra can be found in Fig. 9, where different frequency masks, similar to the previous RF measurement examples, are considered. As it can be seen, the simulations are in good agreement with the measurement results.

The exact complexity numbers corresponding to the full-band and passband linearization scenarios are gathered in Table IV. Such values are calculated for  $U = 1$ ,  $B = 1$ ,  $L = 100$ , and  $N_{\text{OFDM}} = 22528$ . As it can be observed, the proposed beam-level DPD can offer a very significant reduction in the

TABLE IV

DPD MAIN PATH PROCESSING COMPLEXITY PER OFDM SYMBOL AND DPD LEARNING COMPLEXITY PER BLOCK ITERATION FOR THE PROPOSED BEAM-LEVEL DPD FOR SINGLE- AND TWO-USER DIGITAL MIMO WHEN CONSIDERING FULL-BAND/PASSBAND LINEARIZATION. THE COMPLEXITY OF THE PER-TX TIME-DOMAIN DPD IS SHOWN FOR REFERENCE

		Single-user Beam-level FD DPD [MFLOPs]	Two-user Beam-level FD DPD [MFLOPs]	Per-TX TD DPD [MFLOPs]
DPD	IFFT	0.1283	0.2567	—
	FFT	2.1008 / 0.7537	8.403 / 3.0147	—
	BF gen.	0.4964 / 0.1334	1.771 / 0.4173	15.7696
	Filtering	1.0591 / 0.2281	4.4997 / 0.7603	157.6960
	<b>Total</b>	3.7846 / 1.2435	14.9305 / 4.4491	173.4656
Learning	FFT	1.0504 / 0.1284	4.2015 / 0.5135	—
	DPD est.	7.56	302.4	756
	<b>Total</b>	8.6104 / 7.6884	306.6015 / 302.9135	756

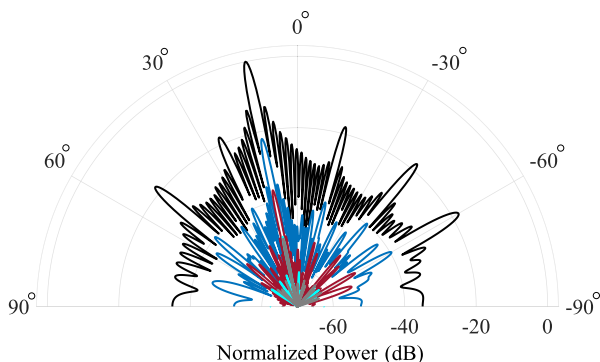


Fig. 10. Far-field beampatterns, with a single dominant user and five low-power users, in terms of total radiated signal (black), total distortion (blue), OOB distortion (red), with multiuser DPD (cyan), and parallel DPD configuration (gray).

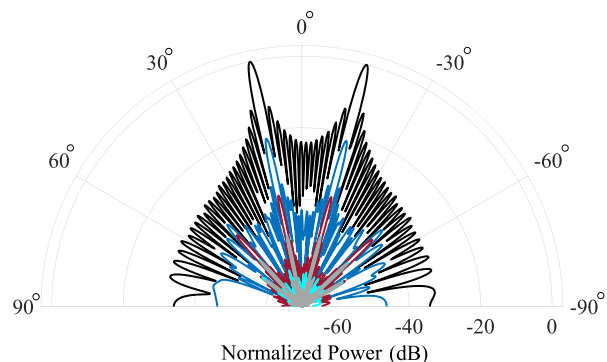


Fig. 11. Far-field beampatterns, with two users served with equal power, in terms of total radiated signal (black), total distortion (blue), total nonlinear distortion (red), with multiuser DPD (cyan), and parallel DPD configuration (gray).

computational complexity. The main reason for this is that the reference per-TX DPD requires  $L = 100$  DPD units running in parallel, whereas the proposed beam-level DPD only needs to take care of linearizing one dominant beam in this example. It is further noted that in the considered simulation scenario, the per-TX DPD and the full-band beam-based DPD entail similar complexity for an array size of for  $L = 2$ . Thus, for any larger antenna array beyond two antenna units and TX chains, the proposed beam-based DPD allows reducing the running complexity of the overall DPD system.

In general, the BS may serve more than one user, resulting in different spatial distortion patterns. However, if one of the users is served with more power than the others, for instance in the case that one user is located at the cell edge, while the rest are closer to the base station, the nonlinear distortion resembles that of the single-user case, and hence, the same set of BFs and associated complexity as in the case considered above is essentially applicable. Example linearization results with one dominant user and five low-power users, considering the full-band DPD and reference per-TX DPD operating in time domain, are shown in terms of the far-field beampatterns in Fig. 10. As it can be seen, excellent linearization can be achieved by considering only SISO BFs and the emissions in the direction of the dominant beam, while in other directions, the emissions are lower. Consequently, also in such cases, the

proposed architecture allows for very significantly reducing the overall complexity of the DPD system. It is further noted that the considered PA models have significant frequency selectivity within the passband, which makes the residual distortion without DPD high. This results in the beampatterns representing the total distortion having around 10 dB higher power than the ones representing the OOB distortion.

To leverage the full potential of the proposed DPD solution, an analysis of the spatial characteristics of the distortion is of paramount importance. Further discussion on this topic is provided in Section VII-D. For clarity, it is noted that, in general, the reference per-TX DPD can provide significantly better linearization outside the directions of the linearized beams because it provides linearization at the PA level, that is, in every spatial direction.

2) *Two-User Case*: When the BS spatially multiplexes several users, intermodulation directions appear, as shown already in Fig. 5(a). In the case of two users being served with the same or similar power, the DPD needs to provide linearization toward four distinct directions, as shown in Fig. 11. This is achieved by following the processing steps detailed in Section IV, which involves considering a multidimensional DPD unit that considers also the different intermodulation BFs. The instantaneous BFs considered for this particular case are  $\Upsilon_{1,1,m}$ ,  $\Upsilon_{2,1,m}$ ,  $\Upsilon_{1,1,m}|\Upsilon_{1,1,m}|^2$ ,  $\Upsilon_{2,1,m}|\Upsilon_{2,1,m}|^2$ ,

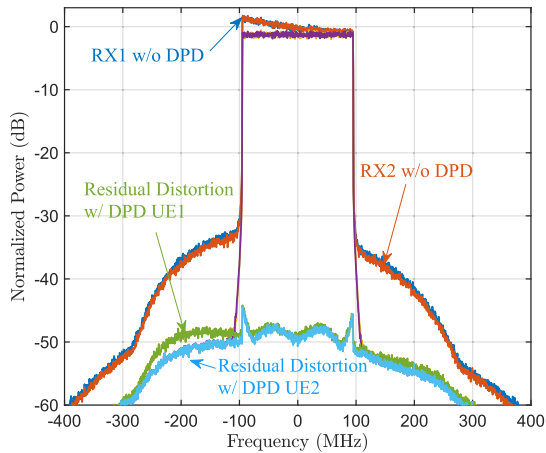


Fig. 12. Spectra of the received signal and residual distortion at both intended receivers with and without the proposed DPD corresponding to the scenario in Fig. 11.

$\Upsilon_{2,1,m}|\Upsilon_{1,1,m}|^2$ ,  $\Upsilon_{1,1,m}|\Upsilon_{2,1,m}|^2$ ,  $\Upsilon_{1,1,m}|\Upsilon_{1,1,m}|^4$ ,  $\Upsilon_{2,1,m}|\Upsilon_{2,1,m}|^4$ ,  $\Upsilon_{1,1,m}^*\Upsilon_{2,1,m}^*$ , and  $\Upsilon_{2,1,m}^*\Upsilon_{1,1,m}^*$ , resulting in  $N_{\text{IBF}} = 10$ , while also two memory taps are considered for each of the BFs, resulting in  $N_{\text{BF}} = 30$ . Furthermore, in the following simulation results, the full-band DPD configuration is considered for presentation simplicity.

The linearization performance, considering the full-band FD DPD and per-TX DPD operating in the time domain, is shown in Fig. 11. As it can be seen, excellent linearization is achieved by the proposed beam-level FD DPD solution toward every considered dominant direction, to an extent similar to the parallel DPD configuration, but entailing ten times less FLOPs. As expected, the per-TX DPD can provide better linearization outside the directions of interest. The corresponding exact complexity values are gathered in Table IV, also showing the complexity of the passband DPD for completeness and reference purposes. In the considered simulation scenario, the per-TX DPD and the full-band beam-based FD DPD entail similar complexity for an array size of  $L = 9$ . Thus, for larger antenna arrays with ten or more antenna units and TX chains, the proposed beam-based FD DPD makes it possible to reduce the running complexity of the overall DPD system.

The corresponding snapshot spectra of the received signals at the two intended receivers are shown in Fig. 12, showing how the nonlinear distortion is effectively reduced. Interestingly, despite considering only the emissions at the main beams and intermodulation directions in the DPD processing, some linearization is also provided in the rest of the directions as it can be observed, e.g., in Fig. 11. The reason for this is that the PA responses, despite being different, are still partly correlated, and hence, some degree of PA-level linearization is achieved. This has been shown through measurement evidence, for instance, in [2, eq. (9)]. This phenomenon can also allow us to reduce the complexity of the beam-level DPD, as the low-power users may not need further linearization through specific DPD processing. It is noted that a further complexity reduction can also be achieved by reducing the linearization BW of the proposed predistorter.

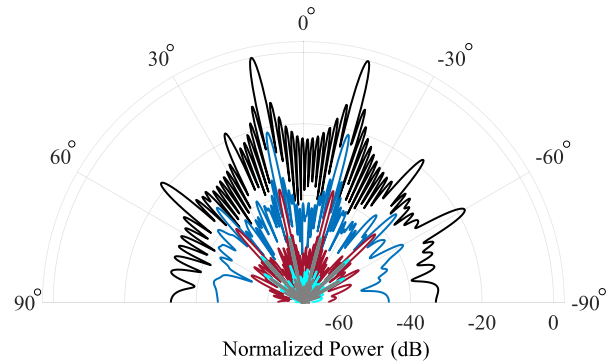


Fig. 13. Far-field beampatterns, with two dominant users and four low-power users, in terms of total radiated signal (black), total distortion (blue), total nonlinear distortion (red), with multiuser DPD (cyan), and parallel DPD configuration (gray).

A similar scenario is shown in Fig. 13, where six users are spatially multiplexed, but there are only two dominant users. In such a case, the characteristics of the distortion resemble those in Fig. 11. Consequently, the same multidimensional BFs as those utilized in the previous evaluation scenario can be used here without entailing further complexity despite serving more users.

### C. Crosstalk Scenario

As a final evaluation scenario, we consider the case in which the TX is affected by PA input and output crosstalk. The baseline crosstalk level is assumed to be at  $-20$  dB between adjacent TX chains and is considered to decay with the square of the distance, implying that the most significant crosstalk originates from adjacent chains. We assume that the BS serves two users with equal power, and hence, the beam-level DPD utilizes the set of BFs indicated in Section VII-B2. The reference parallel DPD structure is considered to employ the SISO BFs, meaning that the crosstalk terms are not modeled.

Performance results in terms of far-field beampatterns are shown in Fig. 14. As it can be observed, the performance of the reference parallel DPD structure is compromised as crosstalk terms are not considered in the DPD model. On the other hand, the proposed beam-level DPD is capable of maintaining the excellent performance shown in the previous evaluation scenarios despite the presence of crosstalk. This is because such architecture is resilient to crosstalk as the reference signal for learning contains a noncoherent sum of the crosstalk signals, while the desired learning signals sum up coherently. Thus, the impact of crosstalk on learning becomes negligible, which is a substantial implementation benefit. Snapshots of the far-field received spectra are also reported in Fig. 15, clearly evidencing the performance loss due to crosstalk with the parallel DPD configuration used as a reference scheme.

### D. Discussion

As it has been shown, the operating principle of the proposed beam-level FD DPD is tightly related to the spatial characteristics of the nonlinear distortion. Hence, proper tracking of the distortion should be considered in the DPD processing.

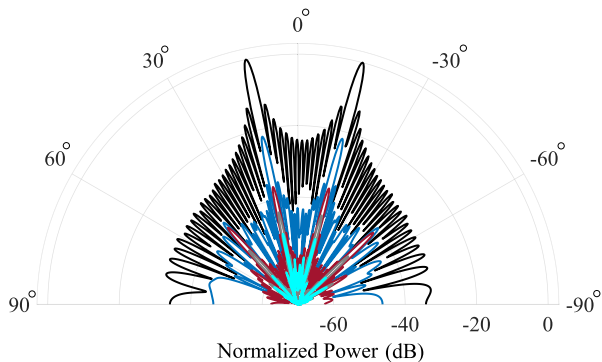


Fig. 14. Far-field beam patterns in the presence of crosstalk, with two users served with equal power, in terms of total radiated signal (black), total distortion (blue), total nonlinear distortion (red), with multiuser DPD (cyan), and parallel DPD configuration (gray).

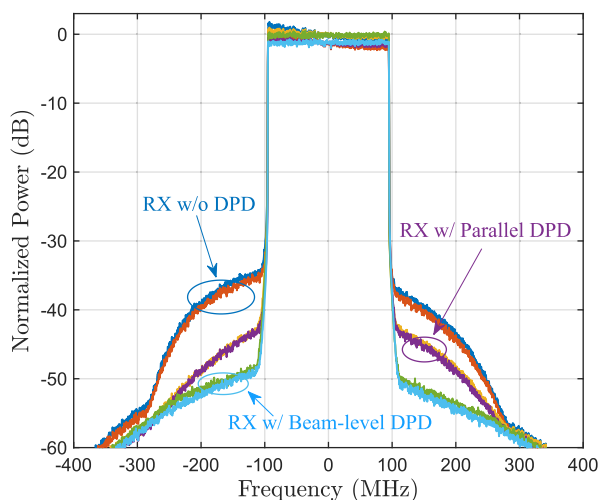


Fig. 15. Spectra of the received signal and residual distortion at both intended receivers, with and without the proposed DPD, in the presence of crosstalk corresponding to the scenario in Fig. 14.

In order to do so, the DPD engine requires information about the beamforming and the relative power levels between the users. Based on our numerical evaluations, a power ratio of at least 14 dB is needed to generate dominant beams. This information can be then exploited to determine which of the different cases we find ourselves in. Obtaining such information is straightforward as the beam directions and the relative powers are deterministic and known at the BS. In addition, the DPD largely benefits from having access to scheduler and link adaptation decisions, e.g., in terms of which data modulations are utilized at different passband frequencies [or physical resource blocks (PRBs)] so that the corresponding DPD order can be tuned to provide sufficient signal quality while keeping the complexity as low as possible.

Despite being less likely, it is also interesting to consider the case in which several users are served with similar power and evaluate the behavior of the distortion in such cases. To that end, the model proposed in [15] can be utilized to assess how much smaller the nonlinear distortion becomes as a result of spatially multiplexing several users. There might be occasions in which the unwanted emissions may already be sufficiently

low that DPD may not be needed. In such situations, the DPD could be turned off until being again required so that the power consumption of the TX can be further reduced.

Finally, as it has been discussed throughout this article, potential victim receivers located at the intermodulation directions are served at the adjacent channels by different BSs. It could also be interesting that the BSs would be aware of the spatial locations of the users that are being served by the neighboring BSs in order to determine whether linearization in such intermodulation directions is required. Investigating such further is part of our future research work.

## VIII. CONCLUSION

In this article, a novel beam-level FD DPD structure has been proposed in order to reduce the overall complexity of DPD systems in array TXs. Its complexity grows proportional to the number of dominant spatially multiplexed users rather than to the number of antennas, which is shown to provide substantial complexity reductions—above an order of magnitude—compared to classical parallel DPD structures. The proposed DPD operates in the FD, which adds an important degree of flexibility so that the DPD operation can be adapted based on instantaneous user quality requirements to enable further complexity reduction. It was also shown that the proposed structure is more robust to potential crosstalk taking place in the array TX, implying that no additional modeling terms are required in order to deal with the crosstalk—opposed to the parallel per-TX DPD structure whose performance is known to be compromised unless higher dimensional DPD models are adopted. Hence, the proposed architecture stands as a promising solution for computationally efficient DPD deployments in fully digital OFDM array TXs. Our future work will include pursuing further hardware-based assessments of the proposed methods while also investigating the base-station coordination aspects when it comes to adjacent channel linearization. Moreover, seeking to formulate and execute piecewise DPD modeling in the FD is an intriguing further opportunity.

## REFERENCES

- [1] M. Abdelaziz, L. Antilla, 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, Jun. 2018.
- [2] A. Brihuega et al., "Piecewise digital predistortion for mmWave active antenna arrays: Algorithms and measurements," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 9, pp. 4000–4017, Sep. 2020.
- [3] X. Liu et al., "Beam-oriented digital predistortion for 5G massive MIMO hybrid beamforming transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 7, pp. 3419–3432, Jul. 2018.
- [4] Q. Luo, X. Zhu, C. Yu, and W. Hong, "Single-receiver over-the-air digital predistortion for massive MIMO transmitters with antenna crosstalk," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 1, pp. 301–315, Jan. 2020.
- [5] X. Wang, Y. Li, C. Yu, W. Hong, and A. Zhu, "Digital predistortion of 5G massive MIMO wireless transmitters based on indirect identification of power amplifier behavior with OTA tests," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 1, pp. 316–328, Jan. 2020.
- [6] 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.

- [7] E. Ng, Y. Beltagy, G. Scarlato, A. B. 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.
- [8] S. Amin, P. Landin, P. Händel, and D. Rönnow, "Behavioral modeling and linearization of crosstalk and memory effects in RF MIMO transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 4, pp. 810–823, Apr. 2014.
- [9] 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, Mar. 2018.
- [10] A. Brihuega, L. Anttila, and M. Valkama, "Neural-network-based digital predistortion for active antenna arrays under load modulation," *IEEE Microw. Wireless Compon. Lett.*, vol. 30, no. 8, pp. 843–846, Aug. 2020.
- [11] A. Brihuega, L. Anttila, and M. Valkama, "Frequency-domain digital predistortion for OFDM," *IEEE Microw. Wireless Compon. Lett.*, vol. 31, no. 6, pp. 816–818, Jun. 2021.
- [12] R. N. Braithwaite, "Amplifier nonlinearities in an antenna array during spatially-multiplexed transmissions," in *Proc. IEEE Radio Wireless Symp. (RWS)*, Jan. 2020, pp. 235–238.
- [13] X. Liu, W. Chen, J. Chu, F. M. Ghannouchi, and Z. Feng, "Multi-stream spatial digital predistortion for fully-connected hybrid beamforming massive MIMO transmitters," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 68, no. 7, pp. 2998–3011, Jul. 2021.
- [14] N. Tervo et al., "Combined sidelobe reduction and omnidirectional linearization of phased array by using tapered power amplifier biasing and digital predistortion," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 9, pp. 4284–4299, Sep. 2021.
- [15] C. Mollén, U. Gustavsson, T. Eriksson, and E. G. Larsson, "Spatial characteristics of distortion radiated from antenna arrays with transmitter nonlinearities," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6663–6679, Oct. 2018.
- [16] 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.
- [17] N. Tervo, J. Aikio, T. Tuovinen, T. Rahkonen, and A. Parssinen, "Digital predistortion of amplitude varying phased array utilising over-the-air combining," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2017, pp. 1165–1168.
- [18] A. Brihuega, L. Anttila, M. Abdelaziz, T. Eriksson, F. Tufvesson, and M. Valkama, "Digital predistortion for multiuser hybrid MIMO at mmWaves," *IEEE Trans. Signal Process.*, vol. 68, pp. 3603–3618, 2020.
- [19] M. Abdelaziz, L. Anttila, and M. Valkama, "Reduced-complexity digital predistortion for massive MIMO," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 6478–6482.
- [20] J. A. Becerra, M. J. M. Ayora, J. Reina-Tosina, and C. Crespo-Cadenas, "Sparse identification of Volterra models for power amplifiers without pseudo-inverse computation," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 11, pp. 4570–4578, Nov. 2020.
- [21] Y. Li, X. Wang, and A. Zhu, "Sampling rate reduction for digital predistortion of broadband RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 3, pp. 1054–1064, Mar. 2020.
- [22] P. P. Campo et al., "Gradient-adaptive spline-interpolated LUT methods for low-complexity digital predistortion," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 68, no. 1, pp. 336–349, Jan. 2021.
- [23] Y. Li, X. Wang, and A. Zhu, "Complexity-reduced model adaptation for digital predistortion of RF power amplifiers with pretraining-based feature extraction," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 3, pp. 1780–1790, Mar. 2021.
- [24] 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.
- [25] 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.
- [26] NR; *Base Station (BS) Radio Transmission and Reception*, document 3GPP TS 38.104, Version 15.9.0, Release 15, Apr. 2020.
- [27] A. A. M. Saleh and R. A. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE J. Sel. Areas Commun.*, vol. SAC-5, no. 2, pp. 128–137, Feb. 1987.
- [28] A. Alkhateeb and R. W. Heath, Jr., "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, May 2016.
- [29] 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.
- [30] 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.
- [31] S. Haykin, *Adaptive Filter Theory*, 5th ed. London, U.K.: Pearson, 2014.
- [32] M. Abdelaziz, L. Anttila, A. Kiayani, and M. Valkama, "Decorrelation-based concurrent digital predistortion with a single feedback path," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 1, pp. 280–293, Jan. 2018.
- [33] M. Z. Waheed et al., "Passive intermodulation in simultaneous transmit-receive systems: Modeling and digital cancellation methods," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 9, pp. 3633–3652, Sep. 2020.
- [34] P. Borwein and T. Erdélyi, *Polynomials and Polynomial Inequalities*. New York, NY, USA: Springer-Verlag, 1995.
- [35] H. Sorensen, M. Heideman, and C. Burrus, "On computing the split-radix FFT," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 1, pp. 152–156, Feb. 1986.
- [36] *Study on Channel Model for Frequencies From 0.5 to 100 GHz*, document 3GPP TR 38.901, Version 15.0.0, Release 15, Jun. 2018.



**Alberto Brihuega** received the B.Sc. and M.Sc. degrees in telecommunications engineering from the Universidad Politécnica de Madrid, Madrid, Spain, in 2015 and 2017, respectively, and the D.Sc.(Tech.) degree (Hons.) in computing and electrical engineering from Tampere University, Tampere, Finland, in 2022.

He is currently an RF System Simulations Engineer with Nokia Mobile Networks, Oulu, Finland. His research interests include statistical and adaptive digital signal processing for compensation of hardware impairments in RF transceivers, RF algorithms, and RF device modeling.



**Lauri Anttila** received the D.Sc. degree (Hons.) in electrical engineering from the Tampere University of Technology (TUT), Tampere, Finland, in 2011.

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 three book chapters. His current research interests include radio communications and signal processing, with a focus on radio implementation challenges in systems, such as 5G, full-duplex radio, and large-scale antenna systems.



**Mikko Valkama** (Fellow, IEEE) 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 with the Communications Systems and Signal Processing Institute, San Diego State University (SDSU), San Diego, CA, USA, as a Visiting Research Fellow. He is currently a Full Professor and the Head of the Unit of Electrical Engineering, Tampere University, Tampere. His general research interests include radio communications, radio localization, and radio-based sensing with particular emphasis on 5G and 6G mobile radio networks.