CHALMERS
UNIVERSITY OF TECHNOLOGY

Over-the-Air Linearization of Phased Array Transmitters Affected by Load Modulation

Downloaded from: https://research.chalmers.se, 2025-02-06 16:26 UTC

Citation for the original published paper (version of record):
Fernandez, J., Anttila, L., Buisman, K. et al (2025). Over-the-Air Linearization of Phased Array Transmitters Affected by Load Modulation. IEEE Transactions on Circuits and Systems I: Regular Papers, In Press. http://dx.doi.org/10.1109/TCSI.2025.3527701

N.B. When citing this work, cite the original published paper.

© 2025 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, or reuse of any copyrighted component of this work in other works.

(article starts on next page)

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Over-the-Air Linearization of Phased Array Transmitters Affected by Load Modulation

Joel Fernandez, *Student Member, IEEE*, Lauri Anttila, *Member, IEEE*, Koen Buisman, *Senior Member, IEEE*, Mikko Heino, *Member, IEEE*, Christian Fager, *Fellow, IEEE*, Thomas Eriksson, *Senior Member, IEEE*, and Mikko Valkama, *Fellow, IEEE*

***Abstract*— Unlocking the potential of millimeter-wave (mmWave) phased array systems demands robust nonlinear transmitter modeling and digital pre-distortion (DPD) techniques. In this article, we present a novel behavioral modeling approach and the corresponding linearization solution for beamforming antenna arrays comprising multiple and mutually interacting nonlinear power amplifier (PA) units. Our non-recursive transmitter model simplifies numerical evaluations across diverse phased array/multiple-input multiple-output (MIMO) configurations under crosstalk-induced load modulation. We introduce a novel, nonlinear forward model parameter identification algorithm tailored for crosstalk-prone array systems and applicable in arbitrary MIMO transmitter configurations, enabling precise modeling and characterization using over-the-air (OTA) observations. Furthermore, we propose an offline direct learning architecture based DPD method, harnessing the estimated nonlinear array forward model and specific beam-sweeping procedure, for linearizing phased arrays under severe load modulation. Numerical assessments across various scenarios demonstrate superior performance, while physical validation on a measurement test bench reinforces our methodology's real-world applicability. Overall, this work paves the way for advanced nonlinear array transmitter optimization and linearization, vital for next-generation wireless communication networks.**

***Index Terms*— 5G, 6G, active array transmitters, antenna crosstalk, digital predistortion, linearization, load modulation, MIMO, phased array, power amplifier, power efficiency.**

I. Introduction

IMPROVING the sustainability and energy-efficiency is among the key criteria and targets, when wireless networks are evolving towards the 6G era [1], [2], [3]. To this end, transmitter (TX) power amplifiers (PAs) constitute a substantial share of the TX power consumption, hence improving the power-efficiency of the PAs and the overall transmitter system is of major importance [4], [5]. Another decisive and related ingredient is to keep the transmit signal passband quality and unwanted emissions within their specified limits [6], [7]. Digital pre-distortion (DPD) technology [8], [9], [10], [11], [12] is among the most established approaches, especially in cellular base-stations and other network transmission points, to mitigate the PA induced nonlinear distortion. Adopting DPD allows pushing the PA units towards their saturation and hence power-efficient operating region, while controlling the antenna signal passband quality and out-of-band and other unwanted emissions. The primary focus of this article is on DPD-based linearization of PA systems, with particular emphasis on modern active antenna array transmitters, while building on our initial work in [13].

Recent advancements in DPD solutions for beamforming arrays [14], [15], [16], [17], [18], [19], [20] focus on tackling the challenge of linearizing a bank of parallel PAs with a single DPD path. A common approach is to consider the linearization along the main-beam direction, allowing to establish an effective single-input single-output (SISO) DPD task, despite differences across the nonlinear characteristics in the involved parallel PA samples [15]. Furthermore, two different principles are commonly considered [14], [19], to provide direct or indirect observations of the outputs of the PAs for DPD parameter learning purposes. These include the conventional *conductive observation method*, where the output signals of individual PAs are measured using directional couplers, phase-aligned in the so-called anti-beamformer, and combined to create the observation signal [14]. As an alternative, *over-the-air (OTA) observation* arrangements can be leveraged and pursued [21]. As the direct conductive observation calls for directional couplers at individual PA level, the research focus is shifting gradually towards the alternative over-the-air observation methods – particularly in millimeter-wave networks where all possible extra insertion losses between the PAs and the antenna units are known to be notably harmful.

Another significant challenge encountered in antenna array based multiple-input multiple-output (MIMO) transmitters stems from the physical *crosstalk* and *other similar interactions among the antennas*. From the linearization and digital predistortion point of view, induced by the involved *load modulation phenomenon*, this makes the exact nonlinear characteristics to be dependent on the applied instantaneous beamforming weights [4]. As a result of such

Received 2 July 2024; revised 4 October 2024; accepted 27 December 2024. This work was supported in part by Business Finland and Vinnova under Grant ENTRY100GHz and in part by the Research Council of Finland under Grant 332361, Grant 338224, and Grant 345654. An earlier version of this paper was presented at the EUSIPCO 2023 [DOI: 10.23919/EUSIPCO58844.2023.10289847]. This article was recommended by Associate Editor J. Walling. *(Corresponding author: Mikko Valkama.)*

Joel Fernandez, Lauri Anttila, Mikko Heino, and Mikko Valkama are with the Department of Electrical Engineering, Tampere University, 33100 Tampere, Finland (e-mail: mikko.valkama@tuni.fi).

Koen Buisman is with the Department of Microtechnology and Nanoscience, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden, and also with the Department of Electrical and Electronic Engineering, University of Surrey, GU2 7XH Surrey, U.K.

Christian Fager is with the Department of Microtechnology and Nanoscience, Chalmers University of Technology, 412 96 Gothenburg, Sweden.

Thomas Eriksson is with the Department of Electrical Engineering, Chalmers University of Technology, 412 96 Gothenburg, Sweden.

Digital Object Identifier 10.1109/TCSI.2025.3527701

© 2025 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

load modulation, multi-antenna systems can suffer from additional nonlinear distortion due to mutual coupling between the different signal paths, which has to be also compensated for [22]. Moreover, within digital array/MIMO transmitter setups, crosstalk can substantially impair the performance of conventional single-input basis function based predistortion methods. Consequently, various advanced multiple-input basis function approaches and related DPD schemes have been studied and assessed [23], [24], [25], [26], [27], [28], [29]. However, such multiple-input nonlinear processing models infer also notable increase in the needed computations, per linearized sample. Specifically, the involved processing and the amount of computations increase significantly as a function of the number of transmitter paths, making them practical only for systems with a limited number of TX paths. Further highlighting this complexity challenge, [30] showed that the power consumption of the linearization system itself can actually surpass that of the power amplifiers when multiple-input DPD methods such as the cross-over DPD (CO-DPD) is adopted. This is particularly so with large waveform bandwidths and growing number of MIMO TX paths. Thus, to address the challenge of incorporating the load modulation effect while controlling the complexity, Hausmair et al. [31] introduced a novel dual-input DPD linearization scheme allowing to effectively capture also the crosstalk effects. This approach enables efficient linearization of the TX with more manageable processing complexity, however, only ordinary per-PA conducted observations are considered in [31].

In this article, building on our early-stage work and results in [13], we address the challenging problems of nonlinear forward model estimation as well as DPD-based linearization of active antenna arrays under crosstalk-induced load modulation via over-the-air observations. We note that the proposed nonlinear forward model identification method is applicable in *arbitrary active array or MIMO systems*, while the developed DPD approach is specifically tailored for *phased-array transmitters* harnessing a dedicated beam-sweeping procedure. Compared to the existing state-of-the-art literature, the main contributions and novelty of the article can be stated as follows:

- We present a comprehensive nonlinear array transmitter model accounting for load modulation between the PA units, applicable across diverse phased array and MIMO transmitter configurations, enhancing the assessment of nonlinear active arrays compared to existing models.
- We introduce a novel nonlinear forward model identification method applicable in arbitrary array/MIMO transmitters operating under crosstalk, in order to characterize the individual PA units and their nonlinearities using OTA observations, while also accounting for the potential memory effects.
- We propose different variants of linear minimum mean squared error (LMMSE) estimators for forward PA model estimation, along with a method to integrate memory into the dual-input polynomial model, thereby enhancing modeling accuracy.
- We introduce a novel strategy for selecting the most optimal set of OTA observation antennas/receivers to improve the efficiency and accuracy of the nonlinear array forward model estimation algorithm.
- We detail and propose a direct learning DPD concept specifically tailored for linearizing millimeter-wave (mmWave) phased array transmitters, harnessing a dedicated beam-sweeping procedure in the parameter estimation stage. The proposed approach utilizes the parameter identification method to obtain forward-path nonlinear array model coefficients, followed by a closed-loop adaptive algorithm for offline learning of the single-input DPD model for phased array linearization.
- Through extensive simulations, we demonstrate the high accuracy of our nonlinear forward model estimation method and the corresponding phased array linearization performance across various system parameters such as the signal-to-noise ratio (SNR) and model identification sample size while considering an 8×8 uniform rectangular array (URA) based phased array TX as a concrete example. Also the impacts of the OTA observation antenna selection and the different LMMSE parameter estimation variants are assessed and shown.
- Finally, as an actual measurement-based proof-of-concept, we validate our proposed methods and concepts on a 4×1 uniform linear array (ULA) test bench using a coupling emulation-based measurement setup, showcasing successful linearization on actual RF hardware, while also addressing the impacts of the nonlinear model order and memory depth.

The outline for the remaining sections of the article is the following. First, Section II provides the fundamental system model for characterizing nonlinear active antenna arrays under load modulation, in terms of complex dual-input polynomials, while also stating the OTA observation model. The proposed nonlinear array forward model identification method is, in turn, formulated in Section III, covering both the basic memoryless nonlinearities as well as nonlinearities with memory. The proposed nonlinear array forward model identification is then applied to the important practical case of phased array linearization in Section IV, in the spirit of direct learning DPD. A wide collection of simulation-based numerical results are next reported and analyzed in Section V, covering both the nonlinear array forward model identification performance as well as the DPD performance in linearizing phased arrays. Complementing the simulation-based assessments, actual RF measurement-based experimental results are then reported and analyzed in Section VI. Finally, Section VII provides the concluding remarks.

II. System Model

The considered array transmitter system scenario is outlined in Fig. 1. Subsequently, following the presentation in our initial work in [13], all models are formulated using complex-valued baseband equivalents in the discrete-time domain. Furthermore, for presentation simplicity, explicit time dependencies are omitted in the basic presentation that focuses on memoryless models. Thus, for instance, the signal $a_{1k}(n)$ is denoted as a_{1k}. Extensions to systems with memory are considered in Sections III and IV. No further assumptions are made about the specific array configuration, and thus the models are valid for all different MIMO/array transmitter systems, such as phased-arrays, hybrid beamforming arrays, and digital MIMO arrays.

A. Models for Nonlinear PA Array Under Crosstalk

1) Recursive Expression for PA Output Signals: Utilizing the dual-input modeling framework from [31], [32], the

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

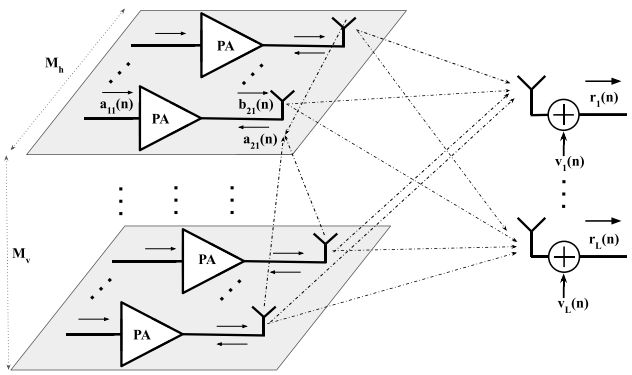

Fig. 1. Illustration of the considered multi-antenna TX system model with K active transmit antennas organized into $M_{\mathrm{h}} \times M_{\mathrm{v}}$ URA architecture, together with L observing antennas. Each radiating path consists of one PA unit connected to an antenna element. For visualization simplicity, coupling signals are depicted only for one PA unit, along with the transmitted signals.

waveform at the power amplifier output in the antenna branch k can be expressed as

$$b_{2k} = \sum_{p=0}^{(P-1)/2} \alpha_{kp} a_{1k}^{(p+1)} a_{1k}^{*\,p} + \sum_{p=0}^{(P-1)/2} \beta_{kp} a_{1k}^{p} a_{1k}^{*\,p} a_{2k} + \sum_{p=1}^{(P-1)/2} \gamma_{kp} a_{1k}^{(p+1)} a_{1k}^{*\,(p-1)} a_{2k}^{*}, \tag{1}$$

where α_{kp}, β_{kp}, and γ_{kp} represent the complex coefficients of the power amplifier model, P denotes the maximum nonlinearity order, while $()^*$ signifies complex conjugation. Referring to the coupling weights as λ_{ki}, the crosstalk waveforms can, in turn, be expressed as

$$a_{2k} = \sum_{i=1, i\neq k}^{K} \lambda_{ki} b_{2i}, \tag{2}$$

where K refers to the total number of PAs and active TX antennas. We note that in practical implementation, the weights λ_{ki} can be acquired, for example, through applicable S-parameter measurements.

Interestingly, it is noteworthy that the expressions in (1) and (2) possess a recursive nature. Hence, we next proceed to derive and provide non-recursive expressions for the PA outputs, which solely rely on the input waveforms. By employing such non-recursive expressions, modeling and characterizing active arrays and their nonlinearities, for example in different simulation environments, can be significantly simplified compared to the recursive expressions in (1) and (2). This constitutes one of the contributions of the article.

2) Non-Recursive Expression for PA Output Signals:

a) Two-antenna case: First, for the basic two-antenna case, we can write individual expressions for b_{21} and b_{22} utilizing (1) and (2) as

$$b_{21} = \sum_{p=0}^{(P-1)/2} \alpha_{1p} a_{11}^{(p+1)} a_{11}^{*\,p} + \sum_{p=0}^{(P-1)/2} \beta_{1p} a_{11}^{p} a_{11}^{*\,p} \lambda_{12} b_{22} + \sum_{p=1}^{(P-1)/2} \gamma_{1p} a_{11}^{(p+1)} a_{11}^{*\,(p-1)} \lambda_{12}^{*} b_{22}^{*}, \tag{3}$$

$$b_{22} = \sum_{p=0}^{(P-1)/2} \alpha_{2p} a_{12}^{(p+1)} a_{12}^{*\,p} + \sum_{p=0}^{(P-1)/2} \beta_{2p} a_{12}^{p} a_{12}^{*\,p} \lambda_{21} b_{21} + \sum_{p=1}^{(P-1)/2} \gamma_{2p} a_{12}^{(p+1)} a_{12}^{*\,(p-1)} \lambda_{21}^{*} b_{21}^{*}. \tag{4}$$

The expressions in (3) and (4), along with their conjugates can be next rewritten in an equivalent matrix form as

$$\begin{bmatrix} 1 & X_{12} & 0 & Y_{12} \\ X_{21} & 1 & Y_{21} & 0 \\ 0 & Y_{12}^{*} & 1 & X_{12}^{*} \\ Y_{21}^{*} & 0 & X_{21}^{*} & 1 \end{bmatrix} \begin{bmatrix} b_{21} \\ b_{22} \\ b_{21}^{*} \\ b_{22}^{*} \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_1^{*} \\ d_2^{*} \end{bmatrix}. \tag{5}$$

This, in turn, can be written in a compact manner as

$$\Delta_{\mathbf{XY}} \mathbf{b} = \mathbf{d}, \tag{6}$$

where $\mathbf{b} \in \mathbb{C}^{(4\times 1)}$ and $\mathbf{d} \in \mathbb{C}^{(4\times 1)}$ read as $\mathbf{b} = \begin{bmatrix} b_{21} & b_{22} & b_{21}^{*} & b_{22}^{*} \end{bmatrix}^T$ and $\mathbf{d} = \begin{bmatrix} d_1 & d_2 & d_1^{*} & d_2^{*} \end{bmatrix}^T$ while $\Delta_{\mathbf{XY}} \in \mathbb{C}^{(4\times 4)}$ is as shown along (5). Furthermore,

$$d_1 = \sum_{p=0}^{(P-1)/2} \alpha_{1p} a_{11}^{(p+1)} a_{11}^{*\,p}, \tag{7}$$

$$d_2 = \sum_{p=0}^{(P-1)/2} \alpha_{2p} a_{12}^{(p+1)} a_{12}^{*\,p}, \tag{8}$$

$$X_{12} = \sum_{p=0}^{(P-1)/2} \beta_{1p} a_{11}^{p} a_{11}^{*\,p} \lambda_{12} b_{22}, \tag{9}$$

$$X_{21} = \sum_{p=0}^{(P-1)/2} \beta_{2p} a_{12}^{p} a_{12}^{*\,p} \lambda_{21}, \tag{10}$$

$$Y_{12} = \sum_{p=1}^{(P-1)/2} \gamma_{1p} a_{11}^{(p+1)} a_{11}^{*\,(p-1)} \lambda_{12}^{*}, \tag{11}$$

$$Y_{21} = \sum_{p=1}^{(P-1)/2} \gamma_{2p} a_{12}^{(p+1)} a_{12}^{*\,(p-1)} \lambda_{21}^{*}. \tag{12}$$

b) K-antenna case: To derive non-recursive models for the output waveforms of the power amplifiers in the more general K-antenna case, we harness the model augmentation principle [33], conceptually similar to (5), in which b_{21} and b_{22}^{*} are stacked together. Switching to vector-matrix notations, we define $\mathbf{b} \in \mathbb{C}^{(2K\times 1)}$ and $\mathbf{d} \in \mathbb{C}^{(2K\times 1)}$ as $\mathbf{b} = \begin{bmatrix} b_{21} & b_{22} & \cdots & b_{2K} & b_{21}^{*} & b_{22}^{*} & \cdots & b_{2K}^{*} \end{bmatrix}^T$ and $\mathbf{d} = \begin{bmatrix} d_1 & d_2 & \cdots & d_K & d_1^{*} & d_2^{*} & \cdots & d_K^{*} \end{bmatrix}^T$ where d_k is given by

$$d_k = \sum_{p=0}^{(P-1)/2} \alpha_{kp} a_{1k}^{(p+1)} a_{1k}^{*\,p}. \tag{13}$$

Now, extending the same formulation as in (6) for the case of K transmit antennas, we define $\Delta_{\mathbf{XY}} \in \mathbb{C}^{(2K\times 2K)}$ as

$$\Delta_{\mathbf{XY}} = \begin{bmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Y}^{*} & \mathbf{X}^{*} \end{bmatrix}, \tag{14}$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

with $\mathbf{X} \in \mathbb{C}^{(K\times K)}$, $\mathbf{Y} \in \mathbb{C}^{(K\times K)}$, and

$$X_{kj} = \begin{cases} 1, & \text{if } k = j \\ -\lambda_{kj} \sum_{p=0}^{(P-1)/2} \beta_{kp} a_{1k}^{p} a_{1k}^{*\,p}, & \text{if } k \neq j \end{cases} \tag{15}$$

and

$$Y_{kj} = \begin{cases} 0, & \text{if } k = j \\ -\lambda_{kj}^{*} \sum_{p=1}^{(P-1)/2} \gamma_{kp} a_{1k}^{(p+1)} a_{1k}^{*\,(p-1)}, & \text{if } k \neq j \end{cases} \tag{16}$$

To further illustrate the structure of $\Delta_{\mathbf{XY}}$, expanding (14) yields

$$\Delta_{\mathbf{XY}} = \begin{bmatrix} 1 & X_{12} & \dots & X_{1K} & 0 & Y_{12} & \dots & Y_{1K} \\ X_{21} & 1 & \dots & X_{2K} & Y_{21} & 0 & \dots & Y_{2K} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{K1} & X_{K2} & \dots & 1 & Y_{K1} & Y_{K2} & \dots & 0 \\ 0 & Y_{12}^{*} & \dots & Y_{1K}^{*} & 1 & X_{12}^{*} & \dots & X_{1K}^{*} \\ Y_{21}^{*} & 0 & \dots & Y_{2K}^{*} & X_{21}^{*} & 1 & \dots & X_{2K}^{*} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_{K1}^{*} & Y_{K2}^{*} & \dots & 0 & X_{K1}^{*} & X_{K2}^{*} & \dots & 1 \end{bmatrix}. \tag{17}$$

A non-recursive expression for $\mathbf{b}$ can then be eventually obtained and formulated as

$$\mathbf{b} = (\Delta_{\mathbf{XY}})^{-1}\mathbf{d}. \tag{18}$$

We note that since $\mathbf{b}$ stacks both b_{2k} and b_{2k}^{*}, only the first K elements are ultimately deployed to obtain the transmitter output. Furthermore, it is important to note that X_{kj} and Y_{kj} are functions of β_{kp} and γ_{kp}, respectively. In the special scenario where load modulation is absent ($\beta_{kp} = 0$ and $\gamma_{kp} = 0$), $X_{kj} = 0$ and $Y_{kj} = 0$ for all $k \neq j$. Consequently, $\Delta_{\mathbf{XY}}$ becomes an identity matrix, and $\mathbf{b} = \mathbf{d}$ in such simplified cases. Therefore, when the influence of load modulation decreases in the system, $\Delta_{\mathbf{XY}}$ tends towards an identity matrix.

B. Model for Over-the-Air Measurements

The presumed model for the over-the-air observation at the l-th observation RX, with $l = 1, 2, \dots, L$, can be formulated as

$$r_l = \sum_{k=1}^{K} \eta_{lk} b_{2k} + v_l \tag{19}$$

where the additive noise v_l is assumed circular complex Gaussian distributed while η_{lk} refers to coupling or channel coefficient from TX element k to observing RX element l. In scenarios where certain elements of the antenna array itself are utilized for observation via OTA coupling, as depicted also later in Fig. 2 within the context of phased array linearization, prior knowledge of the coupling parameters η_{lk} can be presumed. This assumption stems from the fact that the observation receiver antennas are positioned within the same array as the transmitter antennas. In practical implementations, these coefficients can also be acquired via measuring the array's S-parameters, akin to the process of obtaining the coupling coefficients λ_{ki}. We note that the availability and accuracy of such prior knowledge can be a potential practical limitation.

III. Proposed Nonlinear Array Forward Model Identification Method

Harnessing the previous modeling results and our initial work in [13], we next pursue computationally feasible forward model identification methods to characterize an arbitrary active antenna array under load modulation. Notably, the initial results in [13] are extended to the important practical cases of systems with memory, while estimation methods beyond ordinary least-squares (LS) are also pursued and described. Additionally, optimized selection of the observation antenna positions is addressed, and the fundamental processing complexity in terms of the number of floating point operations is quantified.

A. Main Algorithm

Assuming sequences of length N and utilizing (1) as the fundamental basis, we can represent the output of the power amplifier at the k-th antenna branch in vector form as

$$\begin{aligned} \mathbf{b}_{2k} &= [\mathbf{G}^{(0)}(\mathbf{a}_{1k}) \quad \mathbf{G}^{(1)}(\mathbf{a}_{1k}, \mathbf{a}_{2k}) \quad \mathbf{G}^{(2)}(\mathbf{a}_{1k}, \mathbf{a}_{2k})] \\ &\quad \times [\boldsymbol{\alpha}_k^T \quad \boldsymbol{\beta}_k^T \quad \boldsymbol{\gamma}_k^T]^T \\ &= \mathbf{G}(\mathbf{a}_{1k}, \mathbf{a}_{2k})\boldsymbol{\theta}_k \end{aligned} \tag{20}$$

where $\mathbf{a}_{1k}$, $\mathbf{a}_{2k}$, and $\mathbf{b}_{2k}$ all $\in \mathbb{C}^{(N\times 1)}$ and stack N samples of the signals a_{1k}, a_{2k}, and b_{2k}, respectively. The vectors $\boldsymbol{\alpha}_k \in \mathbb{C}^{(Q\times 1)}$, $\boldsymbol{\beta}_k \in \mathbb{C}^{(Q\times 1)}$, $\boldsymbol{\gamma}_k \in \mathbb{C}^{((Q-1)\times 1)}$, and $\boldsymbol{\theta}_k \in \mathbb{C}^{((3Q-1)\times 1)}$, in turn, stack the parameters of the dual-input model with $Q = (P-1)/2 + 1$. Furthermore, the involved matrices $\mathbf{G}(\cdot)$ accommodate the respective nonlinear basis functions, such that $\mathbf{G}^{(0)}(\mathbf{a}_{1k}) \in \mathbb{C}^{(N\times Q)}$, $\mathbf{G}^{(1)}(\mathbf{a}_{1k}, \mathbf{a}_{2k}) \in \mathbb{C}^{(N\times Q)}$, $\mathbf{G}^{(2)}(\mathbf{a}_{1k}, \mathbf{a}_{2k}) \in \mathbb{C}^{(N\times (Q-1))}$, and $\mathbf{G}(\mathbf{a}_{1k}, \mathbf{a}_{2k}) \in \mathbb{C}^{(N\times (3Q-1))}$.

Next, by utilizing the model in (20) together with the measurement model described in (19), we can derive a complete system of equations encompassing two or more observation antennas. This can be written in vector-matrix form as

$$\mathbf{r} = \mathbf{F}\boldsymbol{\theta} + \mathbf{v}, \tag{21}$$

in which

$$\mathbf{F} = \begin{bmatrix} \eta_{11}\mathbf{G}(\mathbf{a}_{11}, \mathbf{a}_{21}) & \dots & \eta_{1K}\mathbf{G}(\mathbf{a}_{1K}, \mathbf{a}_{2K}) \\ \vdots & \ddots & \vdots \\ \eta_{L1}\mathbf{G}(\mathbf{a}_{11}, \mathbf{a}_{21}) & \dots & \eta_{LK}\mathbf{G}(\mathbf{a}_{1K}, \mathbf{a}_{2K}) \end{bmatrix} \tag{22}$$

while $\mathbf{r} = [\mathbf{r}_1{}^T \dots \mathbf{r}_L{}^T]^T$ and $\mathbf{v} = [\mathbf{v}_1{}^T \dots \mathbf{v}_L{}^T]^T$ with both belonging to $\mathbb{C}^{LN\times 1)}$. Furthermore, $\boldsymbol{\theta} = [\boldsymbol{\theta}_1^T \dots \boldsymbol{\theta}_K^T]^T \in \mathbb{C}^{(K(3Q-1)\times 1)}$ and $\mathbf{F} \in \mathbb{C}^{LN\times (K(3Q-1))}$.

Ultimately, as the model in (21) is of linear-in-parameters nature with respect to $\boldsymbol{\theta}$, least-squares based parameter estimation can be pursued. Denoting the pseudo-inverse of $\mathbf{F}$ with $\mathbf{F}^{\dagger}$, this can be formulated as

$$\hat{\boldsymbol{\theta}} = \mathbf{F}^{\dagger}\mathbf{r}. \tag{23}$$

It's important to highlight that a collection of measurements $\mathbf{r}_l, l = 1, 2, \dots, L$, is available through the OTA arrangements. Additionally, samples of the waveforms a_{1k} corresponding to the power amplifier inputs are inherently known. On the other hand, the waveform samples a_{2k} remain *unobservable* directly. Consequently, the estimation of the PA model coefficients

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Algorithm 1 Pseudocode for the Proposed Iterative Approach for Identifying Nonlinear Array Forward Model With OTA Observations. Quantities Associated With $\tilde{()}$ Refer to Local Variables.

Input: a_{1k}, η_{lk}, λ_{ki}, $\mathbf{r}_l$, P and N_D
Initialize the local model outputs as $\tilde{b}_{2k}(0) = 0$
Set the target modeling accuracy NMSE_{des}
Set the initial modeling accuracy $\text{NMSE} = \infty$ and the maximum number of allowed iterations I_{MAX}
while $\text{NMSE} > \text{NMSE}_{\text{des}}$ **or** $i < I_{\text{MAX}}$ **do**
  $i = i + 1$
  Use $\tilde{b}_{2k}(i-1)$ in (2) and compute $\tilde{a}_{2k}(i)$
  Use $\tilde{a}_{2k}(i)$ to construct $\mathbf{F}'$ in (28), to compute $\hat{\boldsymbol{\theta}}'(i)$
  Use $\tilde{a}_{2k}(i)$ and $\hat{\boldsymbol{\theta}}'(i)$ in (25) to compute $\tilde{b}_{2k}(i)$ $\forall\, l$
  Use $\tilde{b}_{2k}(i)$ in (19) to compute all $\tilde{\mathbf{r}}_l(i)$
  Compute $\text{NMSE} = \max(\text{NMSE}(\mathbf{r}_l, \tilde{\mathbf{r}}_l(i)))$
end while
return Nonlinear array forward model $\hat{\boldsymbol{\theta}}' = \hat{\boldsymbol{\theta}}'(i)$

necessitates a recursive approach, owing to the recursive nature and behavior of the expression in (1) already.

Stemming from above, we propose an overall iterative least-squares-based estimation procedure, stated in the form of a pseudocode in Algorithm 1. The presentation in Algorithm 1 covers already the general case of nonlinearities with memory, addressed explicitly in Section III-B. In the proposed approach, the local variables or estimates $\tilde{a}_{2k}$, θ, $\tilde{b}_{2k}$, and $\tilde{r}_l$ are iteratively calculated, seeking to minimize the normalized mean-squared error (NMSE) between the physical observation and its local estimate. Such estimation NMSE, utilized in Algorithm 1, is stated as

$$\text{NMSE}(\mathbf{r}_l, \tilde{\mathbf{r}}_l) = \frac{\sum_{n=1}^{N} |\mathbf{r}_l(n) - \tilde{\mathbf{r}}_l(n)|^2}{\sum_{n=1}^{N} |\mathbf{r}_l(n)|^2}. \tag{24}$$

In general, digital MIMO arrays feature mutually different waveforms in the distinct PA branches. Hence, the system matrix $\mathbf{F}$ is typically of full rank, ensuring a unique solution with any reasonable modulated sequences. However, in the context of phased arrays, addressed in Section IV, the antenna signals stem from the same baseband signal. Consequently, additional measures are necessary to address the resultant rank-deficiency. This is addressed further in Section IV. Furthermore, improved estimators beyond the LS approach are described in Section III-D.

B. Extension to Memory Polynomial Model

We next pursue generalization to systems with memory, through memory polynomials. Notation-wise, we thus bring back the time dependencies that were omitted in Section II, to be able to explicitly characterize the memory effects through sample delays. Thus, for instance, the signal a_{1k} is denoted as $a_{1k}(n)$. We also introduce memory depth, N_D, as a parameter that controls the extent to which the system exhibits memory.

To this end, while now working under the assumption of memory and using the expression in (20) as the starting point, the output of the PA model can now be written in the matrix equivalent form as

$$\mathbf{b}_{2k} = \mathbf{G}'(\mathbf{a}_{1k}, \mathbf{a}_{2k})\boldsymbol{\theta}'_k, \tag{25}$$

where

$$\begin{aligned}\mathbf{G}'(\mathbf{a}_{1k}, \mathbf{a}_{2k}) = [&\mathbf{G}(\mathbf{a}_{1k}(n), \mathbf{a}_{2k}(n)),\\ &\mathbf{G}(\mathbf{a}_{1k}(n-1), \mathbf{a}_{2k}(n-1)), \ldots,\\ &\mathbf{G}(\mathbf{a}_{1k}(n-N_D+1), \mathbf{a}_{2k}(n-N_D+1))],\end{aligned} \tag{26}$$

$$\boldsymbol{\theta}'_k = [\boldsymbol{\theta}_k(0)^T, \boldsymbol{\theta}_k(1)^T, \ldots, \boldsymbol{\theta}_k(N_D-1)^T]^T. \tag{27}$$

In the above formulation, $\mathbf{a}_{1k}(n-k)$ refers to $\mathbf{a}_{1k}(n)$ delayed by k samples. The MP model allows for versatile memory modeling, while the memory depth and model complexity is controlled through N_D. We then build the matrix $\mathbf{F}'$ similarly as in (22), but using $\mathbf{G}'(\mathbf{a}_{1k}, \mathbf{a}_{2k})$ instead of $\mathbf{G}(\mathbf{a}_{1k}, \mathbf{a}_{2k})$. Finally, the complete parameter vector $\boldsymbol{\theta}' = [\boldsymbol{\theta}'^T_1 \ldots \boldsymbol{\theta}'^T_K]^T$ can be estimated with least-squares, as

$$\hat{\boldsymbol{\theta}}' = \mathbf{F}'^{\dagger}\mathbf{r}, \tag{28}$$

in which $\mathbf{F}'^{\dagger}$ denotes the pseudo-inverse of $\mathbf{F}'$.

C. Optimized Selection of Observation Antenna Positions

To perform forward model identification, a certain number of antenna elements from a given TX array can be selected to act as OTA observation receivers. This section introduces a novel and efficient selection algorithm, aiming at choosing those antenna elements that provide the most reliable observations and thus the highest estimation accuracy.

In general, the choice of the observation antennas/receivers is crucial, especially when dealing with a large number of antenna elements and/or a system with high mutual coupling. For instance, if we aim to select $L = 2$ observation receivers out of a 16×1 ULA, there are $\binom{16}{2}$ possible combinations. It is noteworthy that in such systems, the antenna paths close to elements 7 or 8 would experience coupling significantly differently from those at the extreme ends, such as elements 1 or 16. This problem can be further substantiated also for a more generic URA geometry. To this end, consider a scenario where we aim to perform the forward PA model identification of a general phased array system with T antennas, where K elements are transmitting while L elements are acting as OTA observation receivers. The total number of possible placements is given by $C_{\text{tot}} = \binom{T}{L}$, while we denote the corresponding set of the index combinations by Ω_{tot}. Now, inspired by above, the given optimization problem is formulated as a *max-min problem*, written as

$$\Omega_{\text{sel}} = \arg\max_{\Omega_{\text{tot}}} \left[\min_{t=1,\ldots,K} \sum_{n=1}^{L} |\eta_{nt}| \right] \tag{29}$$

where Ω_{sel} refers to the chosen antenna indices. The algorithm follows the 'best of the worst' philosophy, where we determine the placement (observation antenna index combination) by selecting the maximum of the smallest summed coupling from K transmitting antenna elements to a given combination of L observing elements. In Section V, along with the other numerical results, we evaluate and highlight the efficacy of this approach using an 8×8 patch antenna array model, with S-parameters obtained through realistic EM-simulations.

D. Improved Model Identification Through LMMSE Estimation

The least-squares solutions described in (23) and (28) can be improved and generalized to more robust LMMSE estimators.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Depending on the amount of the available side-information, different LMMSE estimator variants are described next. For presentation simplicity, we focus here on the basic case of memoryless nonlinearities while extensions to nonlinearities with memory are straighforward.

1) Full LMMSE Solution: Consider the observation model described in (21). The generalized LMMSE estimator solution for such an observation model can be written as

$$\hat{\boldsymbol{\theta}} = (\mathbf{F}^H \boldsymbol{\Sigma}_{\mathbf{v}}^{-1} \mathbf{F} + \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1})^{-1} \mathbf{F}^H \boldsymbol{\Sigma}_{\mathbf{v}}{}^{-1} \mathbf{r}, \tag{30}$$

where $\boldsymbol{\Sigma}_{\mathbf{v}}$ denotes the covariance matrix of noise, and $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ denotes the covariance matrix of PA coefficients. This serves as the full-performance reference or benchmark estimator, while more practical LMMSE variants are described below.

2) LMMSE I (1^{st} Approximation of Full LMMSE): Assuming first that the noise variables are uncorrelated and of equal variance, i.e., $\boldsymbol{\Sigma}_{\mathbf{v}} = \sigma_v^2 \mathbf{I}$, where σ_v^2 is a scalar quantity denoting the variance of noise and $\mathbf{I}$ is an identity matrix of size $LN \times LN$, (30) can be re-expressed as

$$\hat{\boldsymbol{\theta}} = (\mathbf{F}^H \mathbf{F} + \sigma_v^2 \boldsymbol{\Sigma}_{\theta}^{-1})^{-1} \mathbf{F}^H \mathbf{r}. \tag{31}$$

For practical purposes, $\boldsymbol{\Sigma}_{\theta}$ can then be further assumed or approximated to be a diagonal matrix containing the powers of α, β, and γ across the diagonal.

3) LMMSE II (2^{nd} Approximation of Full LMMSE): Finally, assuming that the PA model coefficients are all of equal variance, i.e., $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \sigma_{\theta}^2 \mathbf{I}$, where σ_{θ}^2 is a scalar denoting the variance of PA model coefficients and $\mathbf{I}$ is an identity matrix of size $K(3Q-1) \times K(3Q-1)$, (31) can be further approximated as

$$\hat{\boldsymbol{\theta}} = (\mathbf{F}^H \mathbf{F} + \frac{\sigma_v^2}{\sigma_{\theta}^2} \mathbf{I})^{-1} \mathbf{F}^H \mathbf{r}. \tag{32}$$

This represents or requires the least amount of side-information in the parameter estimation process, in terms of the second-order statistics of the parameters. The accuracy of the different variants are assessed along the numerical results in Section V.

E. Forward Model Estimation Complexity

In this section, we provide a complexity analysis of Algorithm 1, with particular focus on Steps 7–10 where the actual computations take place. We consider the number of floating point operations (FLOPs) as the main complexity metric, while build directly on the notations that are defined and explained in Sections II and III earlier. Furthermore, in order to derive explicit complexity expressions, a complex multiplication is assumed to cost 6 FLOPs (4 real multiplications and 2 additions) – similar to many other published works [15] seeking to quantify complexity. We also harness the facts that the nonlinear basis functions can be calculated recursively and that delaying basis function samples do not infer any actual FLOPs but can be obtained by simply delaying the samples. Furthermore, the final part of Step 8 involves least-squares estimation, for which the FLOPs count is as described in [34].

Now, assuming the algorithm runs for I_{PA}^{est} iterations, the FLOPs count for Algorithm 1 can be derived and expressed as

$$N_{\text{FLOPs}}^{\text{model}} = I_{\text{PA}}^{est} \Big(N \big(2K \cdot L \cdot N_D (3P+1)^2 + 3K \cdot L \cdot N_D (3P+1) + 8K + 2L(4K-1) + 2N_D(6P+1) + 6P + 14 \big) - \frac{K}{3}(3P+1)^3 \Big) \tag{33}$$

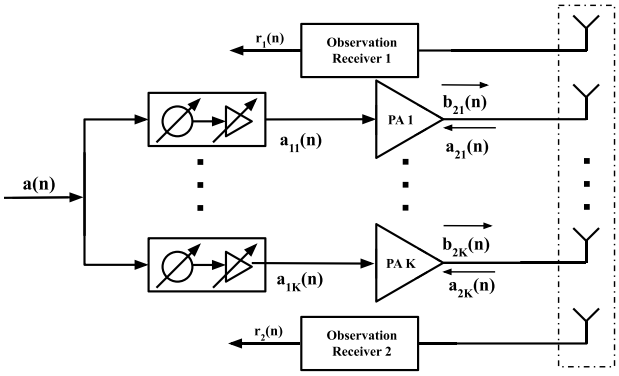

Fig. 2. Conceptual visualization of a uniform linear array with phased-array beamforming where two antenna elements are harnessed for OTA observation purposes in order to identify the involved nonlinearities of the active transmit paths. ULA is considered here only for presentation simplicity while the proposed estimation and linearization concepts are applicable for arbitrary array geometries. Modified and adapted from [13].

As a concrete numerical example, assuming $P = 9$, $N_D = 5$, $L = 2$, $K = 4$, $I_{\text{PA}}^{est} = 10$, and $N = 10\,000$, similar to the actual RF measurements in Section VI, the resulting model estimation complexity over the entire learning period is approximately 6.68 GFLOPs. Such computing counts are generally speaking feasible for modern digital computing platforms, though the more specific further conclusions are subject to the available digital hardware.

IV. Application to Phased Array Linearization

In the subsequent analysis, we direct our attention specifically to millimeter-wave phased arrays and linearizing them using a single DPD path. This involves adopting the proposed forward model estimation, as one element of the overall proposed linearization scheme. Furthermore, we'll assume that some of the antenna elements in the overall antenna array can be deployed for facilitating the over-the-air observations, as shown in Fig. 2.

To this end, for notational clarity, the PA input signals of a beamforming mmWave array with K antenna elements read

$$a_{1k}(n) = w_k a(n), \; k = 1, 2, \ldots, K, \tag{34}$$

where w_k's are the deployed beamforming coefficients, and $a(n)$ denotes the input signal. In what follows, we describe the complete linearization solution harnessing a phased-array beam-sweeping procedure for nonlinear forward model identification, followed by per-beam DPD learning. We build on the initial presentation in [13] while extending to the important case of nonlinearities with memory. Also the linearization processing complexity is quantified through FLOPs.

A. Phased Array Forward Model Identification via Beam-Sweeping

With the PA input signals defined as in (34), the observation model in (21) is not uniquely identifiable due to the rank-deficiency of the involved $\mathbf{F}$. To allow for sufficient rank for model identifiability, we harness a dedicated beamforming procedure over time covering $M_{\text{beam}} \geq K$ discrete angles and

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

mutually independent waveforms. The resulting observations and the corresponding regressor matrices can be then stacked to form aggregated vectors and matrices, respectively, with sufficient rank. The corresponding total sample size is denoted as N_{tot}. The nonlinear forward model identification can then be executed and pursed, via the proposed Algorithm 1.

In practice, such beamformed observations can be acquired during any existing or standardized procedure where the base-station is sweeping its beams over time. One concrete example in 5G context is the transmission of the downlink synchronization signal burst (SSB) [5]. An alternative opportunity is to collect such beamformed observations gradually over time when adjusting the beams between different time slots or subframes for transmitting the actual user data.

B. Forward-Model Empowered Offline Direct Learning for DPD Weights

The overall proposed scheme to determine the DPD weights, empowered by the nonlinear array forward model, is summarized and illustrated in Fig. 3. To this end, we first identify the non-linear model in the forward path, as outlined earlier, utilizing Algorithm 1 while considering the memory polynomial extension described in Section III-B. Then, for the given beam index m, we proceed to learn the actual DPD weights harnessing offline direct learning. This is presented in exact terms next. In general, a single-input DPD is known to allow for efficient linearization of phased arrays, despite antenna coupling and PA load modulation [18]. Therefore, for the main-path DPD processing system, we adopt a single-input memory polynomial model in this work. To this end, we let $a(n)$ to denote the input waveform of the DPD system, while the resulting predistorted waveform can then be expressed as

$$\bar{a}(n) = \sum_{s=0}^{(M_D-1)} \sum_{p=0}^{(P_{\text{dpd}}-1)/2} \beta^m_{\text{dpd},p,s} a^{(p+1)}(n-s) a^{*p}(n-s). \tag{35}$$

In above, $\beta^m_{\text{dpd},p,s}$ reflect the DPD weights, while the subscripts s and p correspond to memory index and nonlinearity order index, respectively. Furthermore, M_D and P_{dpd} denote the memory depth and maximum nonlinearity order of the DPD model, respectively.

In the DPD learning path, the coefficients of the DPD system are calculated and adapted in order to minimize the prevailing error between the linear input $a(n)$ and the locally generated copy of the beamformed far-end signal. For the beam index m, this far-end replica can be expressed as

$$\bar{y}_m(n) = \mathbf{w}_m^H \bar{\mathbf{b}}_{2,m}(n), \tag{36}$$

where $\bar{\mathbf{b}}_{2,m}(n) = [\bar{b}_{21,m}(n) \;\; \bar{b}_{22,m}(n) \ldots \bar{b}_{2K,m}(n)]^T$ while $\mathbf{w}_m^H = [w^*_{1,m} \;\; w^*_{2,m} \ldots w^*_{K,m}]$ denotes the local replica of the LOS channel towards beam direction m. As illustrated also in Fig. 3, the forward models of the involved PA units, identified using the methods described in Section III, are utilized to create the signals $\bar{b}_{2k,m}(n)$. The corresponding error signal, in turn, reads

$$e_m(n) = \frac{\bar{y}_m(n)}{G_m} - a(n), \tag{37}$$

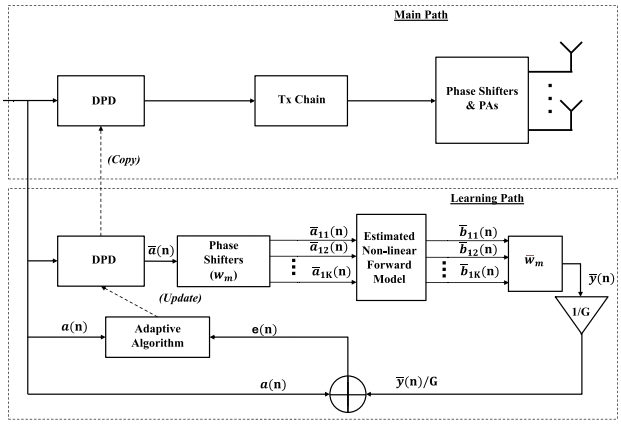

Fig. 3. Block diagram depicting the various processing elements or stages involved in the proposed offline direct learning architecture for phased array linearization. The nonlinear forward model with memory is separately estimated, using Algorithm 1, and is utilized in the learning path to create a local replica of the beamformed signal at intended user direction.

where G_m represents the effective complex linear gain prevailing in the beamformed far-end signal. Following the LS estimation principle, this can be directly estimated through

$$G_m = \frac{\bar{\mathbf{a}}^H \bar{\mathbf{y}}_m}{\bar{\mathbf{a}}^H \bar{\mathbf{a}}}, \tag{38}$$

where $\bar{\mathbf{a}} = [\bar{a}(n) \;\; \bar{a}(n+1) \ldots \bar{a}(n+N_G-1)]^T$, $\bar{\mathbf{y}}_m = [\bar{y}_m(n) \;\; \bar{y}_m(n+1) \ldots \bar{y}_m(n+N_G-1)]^T$, and N_G refers to the sample size in the linear gain estimation process.

The error measure in (37) encapsulates information related to the prevailing nonlinear distortion in the beamformed far-end signal which can be iteratively reduced and minimized using least mean squares (LMS) learning. In this work, we harness block-based learning, combined with orthogonalization for faster and more stable convergence. Defining first the DPD coefficient vector as

$$\boldsymbol{\beta}^m_{\text{dpd}} = [\boldsymbol{\beta}^m_{\text{dpd},0,0} \; \boldsymbol{\beta}^m_{\text{dpd},1,0} \; \cdots \; \boldsymbol{\beta}^m_{\text{dpd},(P_{\text{dpd}}-1)/2,M_D-1}]^T, \tag{39}$$

the corresponding iterative parameter update reads

$$\boldsymbol{\beta}^m_{\text{dpd}}{}^*(i+1) = \boldsymbol{\beta}^m_{\text{dpd}}{}^*(i) - \mu \mathbf{R}^{-1} \mathbf{A}^T(i) \mathbf{e}^*_m(i), \tag{40}$$

where $\mathbf{R} = E[\mathbf{a}(n)\mathbf{a}^H(n)]$ with $\mathbf{a}(n)$ stacking the prevailing instantaneous basis function samples as

$$\mathbf{a}(n) = [a(n) \; a^2(n)a^*(n) \; \ldots \; a^{(P_{\text{dpd}}-1)/2+1}(n-M_D+1)a^{*(P_{\text{dpd}}-1)/2}(n-M_D+1)]^T. \tag{41}$$

Furthermore, the regression matrix $\mathbf{A}(i)$ collects N_{dpd} consecutive instances of $\mathbf{a}(n)$, expressed as

$$\mathbf{A}(i) = \big[\mathbf{a}(n_i) \, \mathbf{a}(n_i+1) \cdots \mathbf{a}(n_i+N_{\text{dpd}}-1)\big]^T, \tag{42}$$

with n_i referring to the time index of the first sample in processing iteration i, while the error vector for beam index m is given as $\mathbf{e}_m(i) = [e_m(n_i) \; e_m(n_i+1) \ldots e_m(n_i+N_{\text{dpd}}-1)]^T$.

C. DPD Processing Complexity

Following a similar methodology and assumptions used for analyzing the complexity of the forward model estimation, we next compute the FLOPs for various stages of the

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

DPD system shown in Fig. 3. For notational convenience, we introduce the total number of DPD basis functions, N_{BF}, expressed as

$$N_{\text{BF}} = M_D \left(\frac{P_{\text{dpd}} - 1}{2} + 1 \right). \tag{43}$$

We then separately analyze the complexity of the DPD main path and the DPD learning path.

1) DPD Main Path: For the DPD forward path or the main path, there are primarily two parts: (1) the generation of the DPD basis functions, and (2) the multiplication of these generated basis functions with the existing DPD coefficients.

Adding together the FLOP counts for these two underlying parts, leads to the following expression for the total FLOPs *per linearized sample* of the DPD main path, given as

$$N_{\text{FLOPs}}^{\text{dpd,main}} = P_{\text{dpd}} + 8N_{\text{BF}}. \tag{44}$$

2) DPD Learning: The DPD learning path entails the functionalities shown in the lower part of Fig. 3. The learning algorithm makes I_{DPD} iterations, processing N_{dpd} samples per iteration. The first step, the DPD block, incurs the same FLOPs per sample per iteration as the DPD block in the main path. In the beamforming step (Phase shifters), there are K complex multiplications, resulting in $6K$ FLOPs per sample per iteration. Running the estimated nonlinear forward model entails basis function generation, followed by multiplication of the basis functions with the PA coefficients. The complexity of these steps was already established in the FLOP calculations for Algorithm 1. We assume recursive calculation of the forward model output with I_{PA}^{run} iterations, following (1)-(2).

In the anti-beamformer step, there are K complex multiplications along with $K-1$ complex additions. The FLOPs for gain estimation and removal are straightforward, the former including one inner product and one squared norm calculation, and the latter entailing one complex multiplication per sample. Finally, the total FLOP count for the block-LMS based adaptive algorithm follows the description in [18]. Overall, the total FLOPs for the entire DPD learning path, over N_{dpd} samples and I_{DPD} learning iterations, can then be expressed as

$$\begin{aligned} N_{\text{FLOPs}}^{\text{dpd,learn}} = I_{\text{DPD}} \Big(& I_{\text{PA}}^{run} \cdot N_{\text{dpd}} \left(8K + 2N_D(6P+1) + 3P + 2\right) \\ & + 6K \cdot N_{\text{dpd}} + 8N_{\text{BF}}^2 + 2N_{\text{dpd}}(4K-1) \\ & + 8N_{\text{dpd}}(N_{\text{BF}}+6) + N_{\text{dpd}}(8N_{\text{BF}} + P_{\text{dpd}}) \\ & + 3(4N_G + 2N_{\text{dpd}} - 1) \Big) + 4N_{\text{dpd}} \end{aligned} \tag{45}$$

Considering similar parametrization to our RF measurements in Section VI, i.e., $P = 9$, $N_D = 5$, $K = 4$, $P_{\text{dpd}} = 9$, $M_D = 5$, $N_{\text{dpd}} = N_G = 10,000$, and $I_{\text{DPD}} = 10$, $I_{\text{PA}}^{run} = 10$ the resulting complexity numbers are 209 FLOPs per sample in the DPD main path and 664 MFLOPs for the DPD learning over the whole learning period. Such computing effort can be concluded very reasonable for any contemporary digital processing system.

V. Numerical Simulation Results

A. Evaluation Tools and Assumptions

To evaluate the performance of the proposed methods, we leverage the dataset of measured PA coefficients available in [18], acquired from Anokiwave PAs operating at 28 GHz, covering both memoryless models as well as models with memory. These measured models correspond to the α coefficients of the more evolved dual-input model in (1), or its variant with memory described in (25)–(27). Subsequently, the synthesis of the remaining dual-input model coefficients, β's and γ's, is achieved by introducing controlled perturbations to α's. More specifically, these coefficients are set at 40 % in strength compared to α's, emulating a severe but realistic load modulation scenario at mmWaves. For presentation simplicity, we consider memoryless models and estimators in this simulations-based section, while the memory-based models and DPD solutions are then deployed in the actual RF measurements-based experiments in Section VI. Furthermore, for notational clarity, we denote the order of the measurement inspired models from [18] by P_{true}, utilized to generate PA output signals in our simulations, while P denotes the order of the estimated model.

For transmit waveform generation, we adopt a 5G NR standard-compatible OFDM waveform featuring a 200 MHz channel bandwidth (BW) and 60 kHz sub-carrier spacing (SCS), corresponding to basic FFT size of $N_{\text{FFT}} = 4096$ [6]. Furthermore, an oversampling factor (OSF) of $N_{\text{os}} = 4$ is utilized to facilitate both PA model identification and DPD processing, alongside the parameter learning. Additionally, we apply additional windowing on top of the baseline OFDM processing to enhance the band-limitation of the ideal digital waveform.

Furthermore, the considered array transmitter setup entails a two-dimensional phased array comprising a total of $T = 64$ antennas with an 8×8 URA geometry and half the wavelength spacing in both horizontal and vertical directions. Among these, $K = 61$ antennas are allocated for transmission, while $L = 3$ antennas are designated for OTA observation purposes. During forward model identification, we consider $M_{\text{beam}} = T = 64$ beams, and assess the performance with $N_{\text{sym}} = \{1, 8, 64\}$ OFDM symbols, as a whole, split equally in time to different considered beams. For generality, the assumed angular range in the beam-sweeping procedure is from $-60°$ to $-60°$, in both horizontal and vertical directions, with uniform angular spacing, and two-dimensional matched filter approach is considered for calculating the beamforming weights. While the value of $N_{\text{sym}} = 64$ implies one full OFDM symbol per beam, we deliberately consider also lower values of $N_{\text{sym}} = 1, 8$, to assess the forward model estimation performance with relatively low sample sizes. To this end, the number of samples in one OFDM symbol period, N_{samp}, reads, in general, $N_{\text{samp}} = N_{\text{FFT}} \times N_{\text{os}}$, and thus the total number of samples, N_{tot}, introduced in Section IV reads $N_{\text{tot}} = N_{\text{FFT}} \times N_{\text{os}} \times N_{\text{sym}}$.

Finally, for coupling modeling in the simulator environment, a realistic patch antenna array was electromagnetically simulated to obtain the accurate coupling model of the considered URA, as analytical coupling models are often inaccurate. Specifically, an 8×8 rectangular linearly polarized patch array was simulated in CST Studio Suite to obtain the S-parameter matrix of the array and thus the coupling matrix between the elements. The element spacing was half of the free-space wavelength, and the patches were simulated on a realistic Rogers RT/duroid 5880 substrate with coaxial feeds to the patches [35].

We next proceed to presenting and analyzing the actual numerical results. For clarity, and importantly, it is highlighted

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

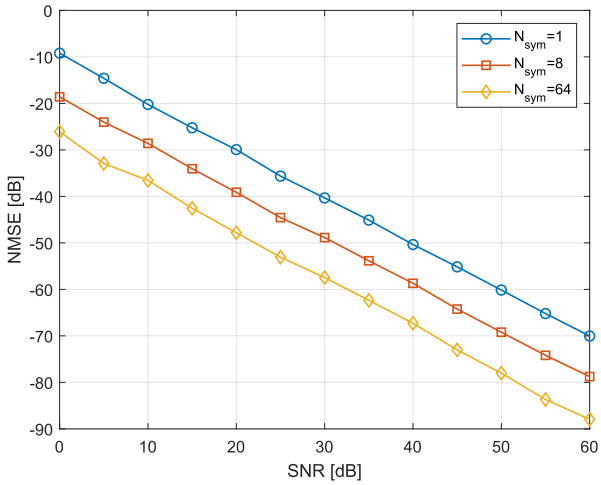

Fig. 4. PA array forward model identification performance with $P = P_{\text{true}} = 9$ and $L = 3$ for varying observation SNR values and for three different signal lengths in the beam-sweeping procedure in terms of the total number of OFDM symbols, N_{sym}.

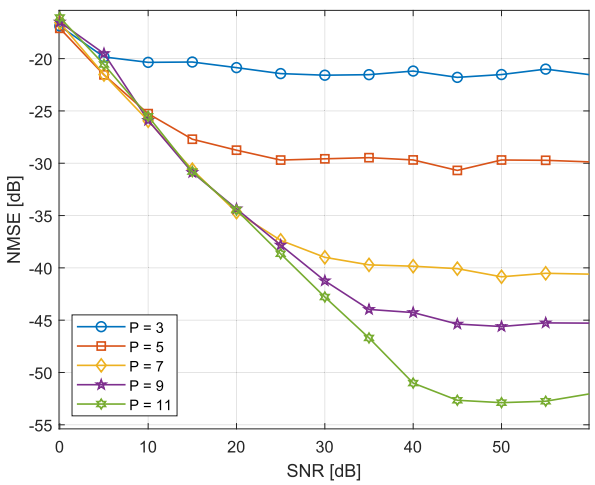

Fig. 5. PA array forward model identification performance with $N_{\text{sym}} = 1$, $L = 3$ and $P_{\text{true}} = 11$ for varying observation SNR values and for five different PA model orders P.

that since the prior-art literature does not contain any existing reference methods that would allow for obtaining forward models for phased arrays under load modulation while observing only through OTA coupling, we focus on assessing the performance of the proposed methods only without explicit further comparisons. The same applies to the actual DPD-based linearization experiments and the corresponding results.

B. Forward Model Identification Performance

1) Performance of the Proposed Algorithm 1 With Least-Squares: Fig. 4 shows the NMSE performance of the proposed PA model identification algorithm (Algorithm 1) for different observation SNR levels, while considering $P_{\text{true}} = P = 9$. Algorithm 1 is here executed for 20 iterations, even though convergence is typically established in fewer iterations already. Furthermore, in these results, the $L = 3$ observation antenna indices are chosen by the proposed selection algorithm in (29). The NMSEs are calculated by averaging over 100 independent Monte Carlo realizations, while also averaging across all the $K = 61$ PAs in the considered array. We can observe from Fig. 4, that we can achieve a better NMSE at a certain SNR by increasing the number of samples of the input training data. Furthermore, increased observation SNR improves the estimation accuracy – however, and importantly, model estimation accuracies in the order of –30 to –40 dB can be reached even when operating at fairly low SNRs in the order of 10 dB. This is an important finding since OTA observations through physical coupling are naturally subject to fairly low or medium SNRs.

Additionally, we study and show the NMSE performance for different PA model orders, P, for fixed $P_{\text{true}} = 11$, while also considering a small sample size with $N_{\text{sym}} = 1$. Otherwise, the assumptions are the same as above. The results shown in Fig. 5 depict the saturating behavior of the modeling accuracy, especially for low model orders P, despite the increasing SNR values. This is understandable since the underlying true nonlinearity is of higher order. On the other hand, if pursuing forward model accuracies in the order of –30 to −35 dB, the results in Fig. 5 also show that such accuracies can be achieved also with relatively low model orders, say $P = 7$, in our setup, assuming SNR values in the range of 15 to 20 dB.

TABLE I

PERFORMANCE COMPARISON OF VARIOUS ESTIMATION ALGORITHMS IN TERMS OF FORWARD MODEL NMSE

SNR [dB]	Estimation NMSE [dB] for different approaches			
	Least-Squares	LMMSE II	LMMSE I	LMMSE (Full)
0	−13.7	−16.8	−19.2	−21.7
5	−17.5	−19.3	−21.2	−23.7
10	−24.1	−25.7	−26.5	−27.8
15	−28.8	−29.9	−32.2	−33.5
20	−34.2	−35.0	−36.7	−38.4
25	−40.1	−41.5	−43.5	−44.3
30	−44.2	−46.0	−47.2	−48.1
35	−49.4	−50.9	−52.3	−53.1
40	−54.1	−56.1	−57.3	−58.0
45	−59.5	−60.4	−61.2	−61.9
50	−64.2	−65.0	−65.9	−66.7
55	−68.9	−69.4	−70.0	−70.5
60	−74.9	−75.0	−75.0	−75.0

2) Performance Comparison of Proposed LMMSE Approaches Against Least-Squares: We next assess the performance of the more elaborate LMMSE-based parameter estimators, proposed in Section III-D, while benchmarking against the baseline least-squares estimation. For presentation simplicity, we consider again the small sample-size case of $N_{\text{sym}} = 1$ and $P_{\text{true}} = P = 7$, with the corresponding results being shown in Table I. We can observe that the different LMMSE variants exhibit notably superior performance compared to the least-squares approach, particularly at low and medium SNR levels – that are of importance when observing through OTA coupling. As a practical concrete example, at an SNR of 15 dB, the difference in the forward modeling accuracy between least-squares and full LMMSE is around 5 dB. In general, we can observe that a given forward modeling accuracy can be obtained at a lower observation SNR when deploying LMMSE estimators, compared to least-squares.

3) Highlighting the Impact of Optimal Observation Antenna Positions: While all the previous results are already building on the proposed observation antenna selection approach in (29), we next concretely showcase and highlight its impact and benefits compared to manually selecting the antenna indices. To this end, Figs. 6 and 7 present visualizations of

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

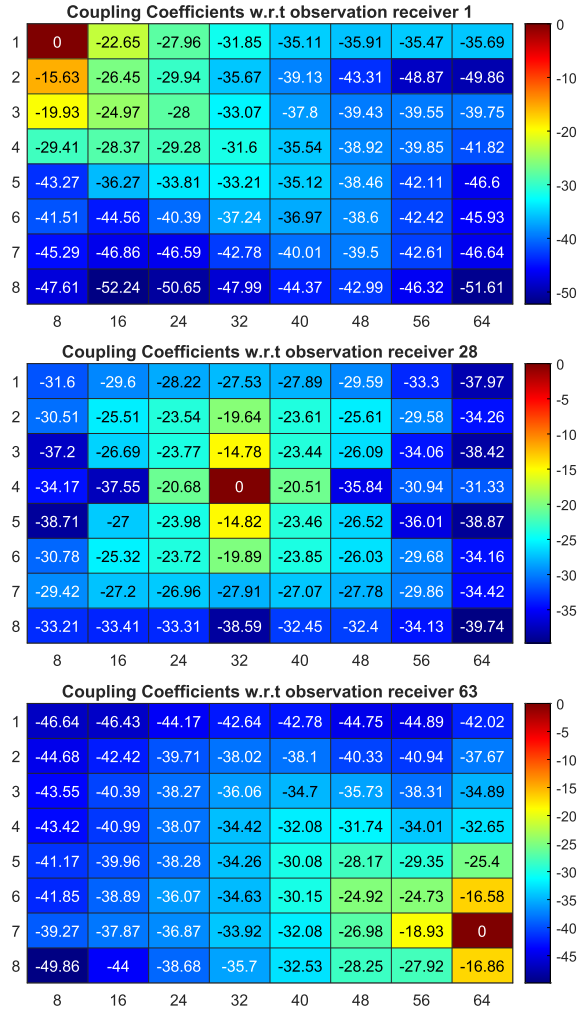

Fig. 6. A colored heatmap visualizing the S-parameters in the considered URA for an example manual selection of observation antennas/receivers as $\Omega_{\text{man}} = \{1, 28, 63\}$.

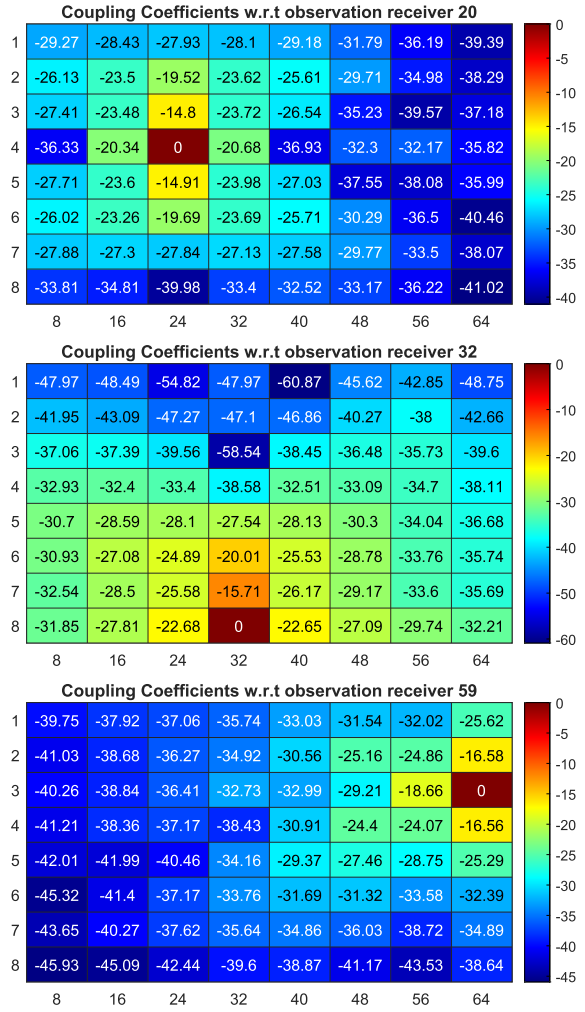

Fig. 7. A colored heatmap visualizing the S-parameters in the considered URA for the set of observation antennas/receivers, $\Omega_{\text{sel}} = \{20, 32, 59\}$, selected by the proposed algorithm in (29).

the S-parameters for a specific antenna element within the considered 8×8 URA, here numbered from 1 to 64. These figures offer a comparison between two scenarios: one where observation antennas are manually selected and another where they are chosen by the algorithm outlined in Section III, Equation (29). Notably, the algorithmically selected set of $L = 3$ observation antennas out of $T = 64$ demonstrates enhanced coverage over all antenna elements of the URA. These improvements are further corroborated by the results in Fig. 8, which showcases the NMSE performance for forward PA model identification across various placements of OTA observation antennas, over different estimation iterations in Algorithm 1, in the example case of $N_{\text{sym}} = 1$ and $P_{\text{true}} = P = 5$. Furthermore, for visualization purposes, we deliberately set the SNR to a high value of 40 dB. Importantly, these results clearly confirm the advantage of using the observation antenna set selected by the algorithm as to reach desired NMSE level in the least number of iterations.

C. DPD Linearization Performance

We next proceed to evaluate the actual linearization performance of the proposed DPD approach, harnessing the overall proposed offline direct learning approach shown in Fig. 3. As a concrete example case study, we consider $N_{\text{sym}} = 1$ and $P_{\text{true}} = P = 5$ in the forward model estimation phase while the actual DPD stage considers a DPD model of order $P_{\text{dpd}} = 9$, maintaining consistency with the previously described array transmitter scenario. Within the closed-loop learning system with gradient-descent DPD learning algorithm, we consider a step-size of $\mu = 0.0003$. Furthermore, to study the impact of the forward model accuracy on the actual linearization performance, we allow varying observation SNRs such that different forward model accuracies can be assessed as part of the overall DPD system.

Firstly, Fig. 9 demonstrates the linearization effectiveness of the proposed DPD system by comparing the normalized power spectral densities (PSDs) of the received signal at the far-end user, both with and without DPD. This example illustration is calculated for an example transmit beam pointing towards 0° (horizontal angle), –30° (vertical angle), while also allowing for and illustrating the impacts of different forward model identification NMSEs. Table II then gathers and illustrates the actual EVM and the adjacent channel power ratio (ACPR) figures of merit of the DPD system across various levels of forward model identification NMSEs, here being also averaged across all different beamforming angles.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

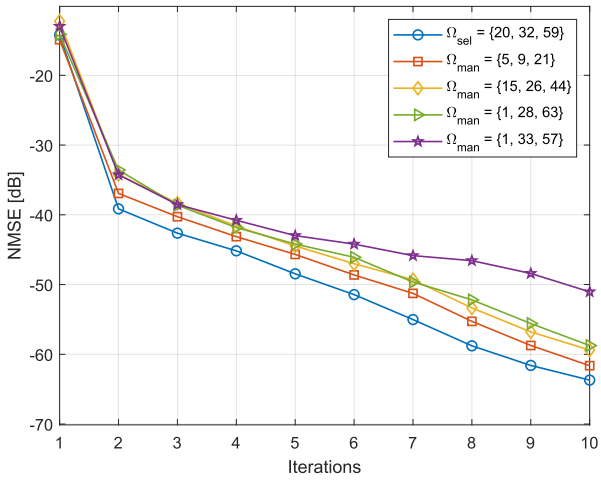

Fig. 8. Forward model NMSE performance over the Algorithm 1 iteration counts, comparing different possible placements for OTA observation antennas. Here, Ω_{sel} is as obtained through the algorithm in (29) while Ω_{man} refers to the set of observation receivers selected manually as shown in the legend.

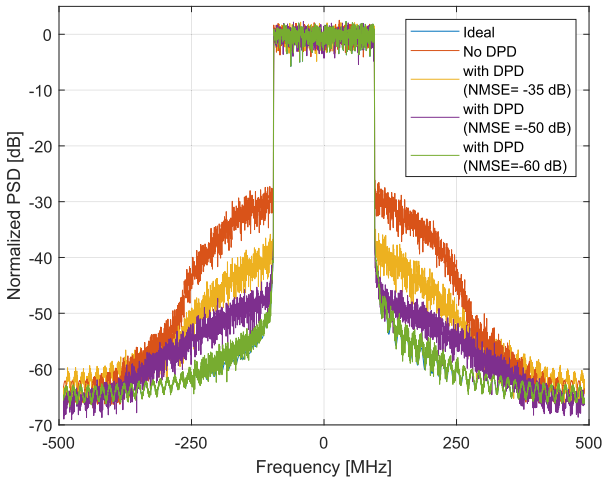

Fig. 9. DPD linearization example showing the beamformed spectra at far-end user without and with DPD, for various PA model estimation NMSEs. $P_{\text{true}} = P = 5$ and $P_{\text{dpd}} = 9$.

We can clearly observe the good linearization performance already at forward modeling NMSEs of –35 dB to –40 dB, while further enhanced linearity can be achieved when the forward model accuracy is higher.

VI. Experimental RF Measurement Results

A. Measurement Setup

Due to the lack of actual array hardware with embedded OTA observation capabilities, measurements were performed using the RF WebLab platform [36], which is an online system for power amplifier testing with active load pull capabilities. The system includes two Vector Signal Transceivers (VST) NI PXIe-5646R and NI PXIe-5645R operating at 2.14 GHz, one of which is limited to a bandwidth of 100 MHz. The transmitters of the VSTs create the input signal and the load pull signal at the output of the device under test (DUT). Additional components include custom driver amplifiers protected by isolators and directional couplers at input and output to measure the required calibrated power waves using the receivers of the VSTs. The system is calibrated using short, open, and load standards as well as a power meter at reference planes at the DUT's connectors. The device under test was a 6W GaN power amplifier (CGH40006TB) on a testing and evaluation board.

Furthermore, employing the coupling emulation technique presented in [37], we are iteratively able to introduce controlled coupling corresponding to any given PA array configuration. For these experimental validations, we focus on a simple 6×1 ULA as an elementary proof-of-concept, with $K = 4$ active antennas, $L = 2$ observing antennas, and half the wavelength spacing between the neighboring elements. For simplicity, the antennas #1 and #6 are used for observing while the others for transmitting and radiating. Despite the fairly simple or straightforward ULA arrangements, this setup captures and supports the fundamentals of the work and the proposed methods – and the overall underlying problem setting: linearizing an effective array of PAs subject to load modulation, while observing through OTA coupling.

Finally, as the digital baseband waveform, we utilize a 5G NR standard-compliant OFDM waveform corresponding to 30 MHz channel bandwidth and 30 kHz sub-carrier spacing (natural FFT size of 1024) [6], alongside an oversampling factor of 8, $N_{\text{sym}} = 10$, and time-domain windowing for sidelobe suppression. The reduced channel bandwidth, compared to the previous simulations-based results, is primarily due to the bandwidth limitations of the available hardware and test bench.

TABLE II

Linearization Performance of the Proposed DPD Approach for Different NMSEs in PA Array Forward Model Estimation, $P_{\text{TRUE}} = P = 5$ and $P_{\text{DPD}} = 9$

NMSE of PA model estimation	Simulation Results			
	EVM (w/o DPD)	EVM (with DPD)	ACPR (w/o DPD)	ACPR (with DPD)
−35 dB	4.21%	2.64%	−34.89 dB	−45.02 dB
−40 dB	4.21%	2.54%	−34.89 dB	−48.80 dB
−45 dB	4.21%	2.34%	−34.89 dB	−49.34 dB
−50 dB	4.21%	2.11%	−34.89 dB	−51.10 dB
−55 dB	4.21%	1.85%	−34.89 dB	−55.80 dB
−60 dB	4.21%	1.78%	−34.89 dB	−61.60 dB

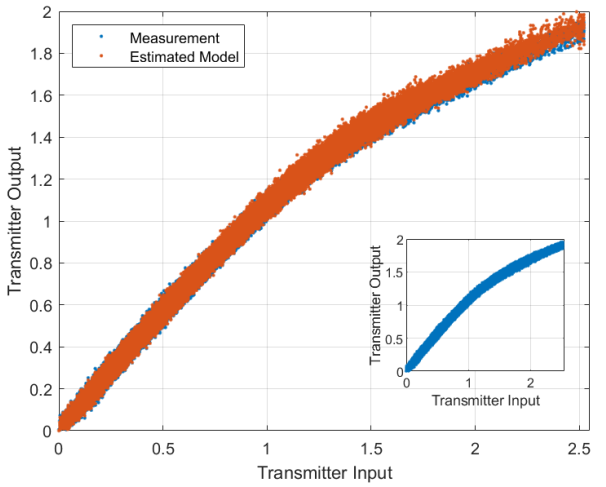

Fig. 10. AM-AM plots depicting the fitting of the response of the estimated forward model to the measured response for a single example PA entity of the considered array arrangement under load modulation. Here, $P = 9$ and $N_D = 5$.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

B. Measured Forward Modeling Accuracy

In this subsection, we first assess the forward modeling accuracy of the Algorithm 1 presented in Section III through the actual measurements. Different from the previous simulations, here the actual *memory-based models* are utilized – in general for both forward modeling as well as for actual DPD processing. To this end, Fig. 10 first shows and illustrates the AM-AM plot of one of the PAs in the considered ULA arrangement, comparing the response of the measurement as well as the estimated forward model when considering the forward model order of $P = 9$ and memory depth of $N_D = 5$. Furthermore, Fig. 11 shows and quantifies the measured NMSE performance of forward model estimation while varying the considered polynomial order (P) and memory depth (N_D). We can observe that for this experimental measurement setup, the NMSE converges after $P = 5$ and $N_D = 4$. Furthermore, we can observe that forward modeling accuracies up to $-40 \ldots -41$ dB can be achieved in these measurements, via the proposed methods. These identified forward model coefficients will be next utilized, as part of the overall proposed DPD framework, for linearization purposes.

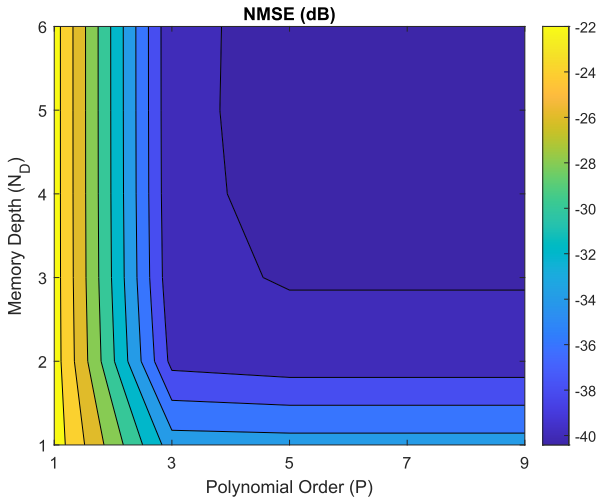

Fig. 11. Filled 2-D contour plot showing the measured forward model estimation performance for the considered experimental PA array arrangement under load modulation, in terms of NMSE, for varying polynomial order and memory depth.

C. Simulated DPD Performance With Measured Forward Model

We proceed next towards assessing the linearization performance of the proposed DPD approach, with the measured forward model. Specifically, we utilize a forward model of order $P = 9$ and of memory depth $N_D = 5$ while the DPD model is of order $P_{\text{dpd}} = 13$, consistent with the array transmitter scenario described earlier. Furthermore, the memory depth for the DPD model is $M_D = 3$. A back-off of -3 dB is applied with respect to the input signal to facilitate the linearization process. Within the closed-loop system, we set the step-size of the gradient-descent learning algorithm to $\mu = 0.0002$.

Initially, before the final assessment of truly measured DPD performance, we demonstrate the effectiveness of the proposed DPD system in linearizing the measured PA model in the simulator environment. That is, the measured forward model is ported into the previous simulation environment, for initial linearization testing and assessments. To this end, the normalized PSDs, both with and without DPD, are depicted and shown in Fig. 12, calculated at the far-end user while considering an example beamforming angle of –30°. We can observe that the linearization of the measured PA array forward model works well, with the ACPR tending towards –50 dB range. Corresponding quantitative results, covering also different beamforming angles, are available and shown in Table III.

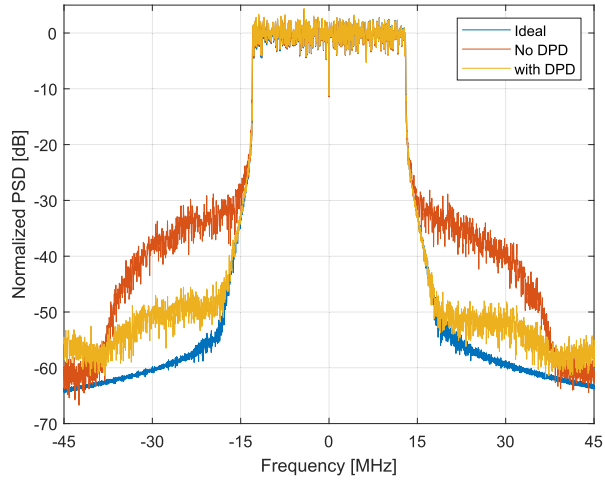

Fig. 12. PSD plot depicting the linearization performance of the proposed DPD approach in linearizing the measured PA model array under load modulation in the simulator environment.

D. Measured DPD Performance

Finally, using the same setup consisting of a forward model of order $P = 9$ and memory depth of $N_D = 5$, and a DPD model of order $P_{\text{dpd}} = 13$ and memory depth of $M_D = 3$, we test and demonstrate the experimental linearization performance of the proposed DPD scheme. To this end, Fig. 13 depicts the measured linearization performance by showing and comparing the normalized PSDs of the received signal at the far-end user, both with and without DPD, for an example beamforming angle of –30°. The corresponding quantitative results, in terms of EVM and ACPR, are gathered and shown along Table III, covering also different alternative beamforming angles. We can observe that the measured linearization performance is systematically in the order of -47 dB to -50 dB, independent of the beamforming angle, evidencing high-quality linearization despite the strong load modulation effects in the experimental system.

Finally, we assess the impact of the considered backoff level utilized in DPD learning and linearization experiments. The obtained results with backoff ranging from 3 dB down to 0.5 dB are shown in Table IV, highlighting that ACPR gains in the order of 10 dB can still be obtained even if a small backoff value of 1 dB is utilized. However, to reach ACPRs in the order of –45 dB, appropriate backoff region can be observed to be 1.5 dB and above. One of the future research topics thus covers developing DPD learning and processing models that allow for reduced backoff values, further improving the system power efficiency.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE III

PERFORMANCE OF THE PROPOSED DPD METHODS IN LINEARIZING A 4 × 1 PA ARRAY AT 3 DB BACKOFF FOR DIFFERENT BEAMFORMING ANGLES. COMPARING LINEARIZATION OF A MEASURED ARRAY MODEL IN A SIMULATION ENVIRONMENT AND ACTUAL MEASURED LINEARIZATION RESULTS

Beamforming angle (in degrees)	DPD simulation with a measured PA array model				DPD measurement with an actual PA array experiment			
	EVM (w/o DPD)	EVM (with DPD)	ACPR (w/o DPD)	ACPR (with DPD)	EVM (w/o DPD)	EVM (with DPD)	ACPR (w/o DPD)	ACPR (with DPD)
−60	5.21%	3.05%	−35.82 dB	−52.29 dB	5.43%	3.00%	−35.37 dB	−50.07 dB
−30	4.57%	2.42%	−36.00 dB	−53.28 dB	5.07%	2.74%	−35.29 dB	−50.36 dB
0	5.04%	3.01%	−35.63 dB	−50.01 dB	5.15%	2.79%	−35.50 dB	−46.90 dB
30	4.59%	2.61%	−36.55 dB	−51.25 dB	5.18%	2.78%	−35.51 dB	−49.60 dB
60	4.96%	2.85%	−36.57 dB	−51.15 dB	5.91%	3.50%	−35.14 dB	−49.65 dB

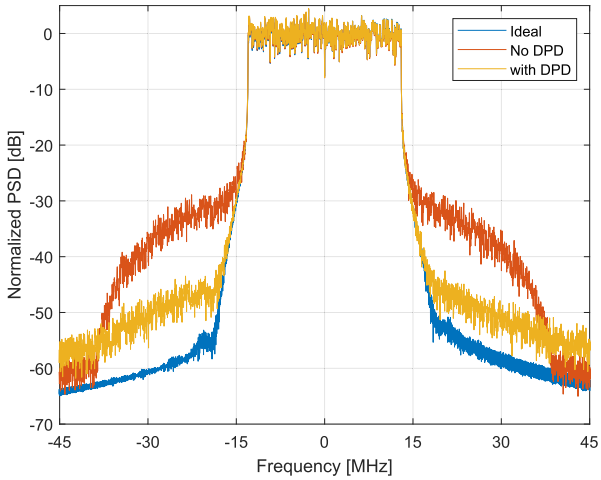

Fig. 13. PSD plot depicting the linearization performance of the proposed DPD approach in linearizing PA array under load modulation in the actual experimental environment.

E. Discussion and Potential Practical Limitations

Like already noted, one potential practical limitation in the proposed methods is the availability and accuracy of the assumed S-parameter characteristics for the involved active array. Additionally, several factors beyond the potential modeling and parameter learning limitations may contribute to practical limitations, including the following aspects.

1) Dynamic Range: The linearization performance observed through the measured model in a simulator environment may not translate completely to the actual measured linearization capability due to differences in dynamic range. The model may operate under some assumptions or constraints that do not accurately define the operating conditions of a real-world PA system. This can lead to differences in performance, especially in situations where the PA operates in or near its saturation region.

2) Environmental Factors: Different physical environment related factors such as temperature variations [38], power supply variations and load imbalances can affect the behavior of PA systems over time. These factors may not be adequately considered in the model, leading to a potential performance reduction. However, these can be remedied through updating the nonlinear forward model sufficiently often, while then also adapting the actual DPD weights correspondingly.

3) Implementation Challenges: Effective use of DPD algorithms through actual real-time digital hardware introduces additional challenges that cannot be fully captured in the model. This includes factors such as quantization effects, finite precision arithmetic, processing delay, and synchronization, which can all affect the efficiency of linearization methods when implemented on real hardware. In our case, however, the DPD algorithms are computed and executed in floating point arithmetic, hence only the possible residual errors in time and/or frequency synchronization between the transmit path and observation receiver paths may have contributed to the measured linearization performance.

TABLE IV

MEASURED LINEARIZATION PERFORMANCE OF THE PROPOSED DPD SOLUTION AT DIFFERENT BACK-OFF LEVELS

Back-off [dB]	EVM		ACPR	
	w/o DPD	with DPD	w/o DPD	with DPD
0.5	7.54%	4.82%	−30.67 dB	−36.14 dB
1.0	7.23%	4.16%	−31.61 dB	−39.36 dB
1.5	6.57%	3.05%	−32.67 dB	−43.18 dB
2.0	5.73%	2.90%	−33.70 dB	−46.23 dB
2.5	5.29%	2.88%	−33.97 dB	−47.26 dB
3.0	5.18%	2.78%	−35.53 dB	−49.67 dB

VII. CONCLUSION

In this article, we have addressed the challenging problems of forward modeling and DPD-based linearization of nonlinear array transmitter systems under load modulation while observing only in over-the-air manner through physical coupling. First, a novel forward modeling and parameter identification approach was proposed, building on dual-input behavioral models and iterative least-squares-based estimation processing while also accommodating memory in the system. Different variants of more elaborate LMMSE estimators were also described, beyond ordinary least-squares, while also an efficient max-min optimized problem was formulated for obtaining optimum observation antenna locations within the overall considered array. All the developed nonlinear forward modeling and model identification methods are applicable in arbitrary array transmitters, independent of the applied MIMO or beamforming scheme. Secondly, DPD-based linearization of phased arrays under load modulation was pursued, while operating under OTA observations only. A direct-learning based architecture utilizing the dual-input nonlinear forward model, estimated using the proposed methods, was described, together with computationally efficient gradient-descent parameter learning methods for obtaining the DPD coefficients. The DPD method is specifically tailored for and thus primarily applicable in phased-array transmitters, harnessing a specific beam-sweeping procedure. Through extensive simulations as well as physical validation on an RF measurement test-bench, the capabilities and performance

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

advantages of the various proposed methods were assessed and demonstrated, shown to facilitate highly accurate forward modeling as well as DPD-based linearization across various scenarios.

Our future work will focus on extending the actual DPD-based linearization solutions to digital MIMO transmitters while also improving the current methods to operate efficiently under lower input backoff values. Additionally, as the current proposed methods rely on S-parameter measurements for the array coupling coefficients – which can be seen as one practical limitation – our future work will also consider developing machine learning aided linearization solutions agnostic to the knowledge of the underlying coupling phenomenon.

REFERENCES

[1] *Framework and Overall Objectives of the Future Development of IMT for 2030 and Beyond*, document 2160-0, ITU-R, 2023.

[2] C.-X. Wang et al., "On the road to 6G: Visions, requirements, key technologies, and testbeds," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 2, pp. 905–974, 2nd Quart. 2023.

[3] H. Tataria, M. Shafi, A. F. Molisch, M. Dohler, H. Sjöland, and F. Tufvesson, "6G wireless systems: Vision, requirements, challenges, insights, and opportunities," *Proc. IEEE*, vol. 109, no. 7, pp. 1166–1199, Jul. 2021,

[4] 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.

[5] H. Holma, A. Toskala, and T. Nakamura, *5G Technology: 3GPP New Radio*. Hoboken, NJ, USA: Wiley, 2019.

[6] *NR; Base Station (BS) Radio Transmission and Reception (Release 18)*, Standard TS 38.104, 3GPP, Apr. 2023.

[7] K. Chuang, H. Yektaii, N. Outaleb, A. Raslan, S. Bhal, and P. Forbes, "Towards sustainable networks: Attacking energy consumption in wireless infrastructure with novel technologies," *IEEE Microw. Mag.*, vol. 24, no. 12, pp. 44–59, Dec. 2023.

[8] 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.

[9] A. Fischer-Bühner et al., "Sparsely gated mixture of experts neural network for linearization of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 72, no. 7, pp. 4367–4382, Jul. 2024.

[10] 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.

[11] M. Othmani, N. Boulejfen, M. Turunen, M. Allén, F. M. Ghannouchi, and M. Valkama, "Parallel delta-sigma modulator-based digital predistortion of wideband RF power amplifiers," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 70, no. 2, pp. 705–718, Feb. 2023.

[12] C.-F. Cheang, P.-I. Mak, and R. P. Martins, "A hardware-efficient feedback polynomial topology for DPD linearization of power amplifiers: Theory and FPGA validation," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 65, no. 9, pp. 2889–2902, Sep. 2018.

[13] J. Fernandez, L. Anttila, M. Valkama, and T. Eriksson, "Linearization of active array transmitters under crosstalk via over-the-air observations," in *Proc. 31st Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2023, pp. 1420–1424.

[14] 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, Jun. 2018.

[15] 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.

[16] 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.

[17] 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.

[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] N. Tervo et al., "Digital predistortion of phased-array transmitter with shared feedback and far-field calibration," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 1, pp. 1000–1015, Jan. 2021.

[20] 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.

[21] K. Hausmair, U. Gustavsson, C. Fager, and T. Eriksson, "Modeling and linearization of multi-antenna transmitters using over-the-air measurements," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2018, pp. 1–4.

[22] M. Romier, A. Barka, H. Aubert, J.-P. Martinaud, and M. Soiron, "Load-pull effect on radiation characteristics of active antennas," *IEEE Antennas Wireless Propag. Lett.*, vol. 7, pp. 550–552, 2008.

[23] X. Wu, X. Pan, P. Sun, and J. Zhou, "Scalable double crosstalk canceling digital predistortion for MIMO transmitters," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 71, no. 12, pp. 6292–6305, Dec. 2024.

[24] 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 Techn.*, vol. 62, no. 4, pp. 810–823, Apr. 2014.

[25] 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.

[26] 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.

[27] Shipra, G. C. Tripathi, and M. Rawat, "Power amplifier linearization in the presence of crosstalk and measurement noise in MIMO system," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 69, no. 10, pp. 3988–3992, Oct. 2022.

[28] T. Uthayakumar, A. Abdelhafiz, X. Wang, and M. Jian, "Scalable single-input behavioral modeling architecture for MIMO systems with crosstalk," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 70, no. 5, pp. 2179–2188, May 2023.

[29] 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.

[30] J. Zanen, E. Klumperink, and B. Nauta, "Power efficiency model for MIMO transmitters including memory polynomial digital predistortion," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 68, no. 4, pp. 1183–1187, Apr. 2021.

[31] 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.

[32] C. Fager et al., "Prediction of smart antenna transmitter characteristics using a new behavioral modeling approach," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jul. 2014, pp. 1–4.

[33] T. Adali, P. J. Schreier, and L. L. Scharf, "Complex-valued signal processing: The proper way to deal with impropriety," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5101–5125, Nov. 2011.

[34] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3, Ed., Baltimore, MD, USA: The Johns Hopkins Univ. Press, 1996.

[35] Rogers Corp. *RT/duroid 5870/5880 Datasheet*. Accessed: Aug. 2023. [Online]. Available: https://www.rogerscorp.com/advanced-electronics-solutions/rt-duroid-laminates/rt-duroid-5880-laminates

[36] RF WebLab. *Chalmers University of Technology*. Accessed: Feb. 2024. [Online]. Available: http://dpdcompetition.com/rfweblab/

[37] D. Nopchinda and K. Buisman, "Measurement technique to emulate signal coupling between power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 4, pp. 2034–2046, Apr. 2018.

[38] G. Jindal, G. T. Watkins, K. Morris, and T. A. Cappello, "Digital predistortion of RF power amplifiers robust to a wide temperature range and varying peak-to-average ratio signals," *IEEE Trans. Microw. Theory Techn.*, vol. 70, no. 7, pp. 3675–3687, Jul. 2022.