Training Data Selection and Dimensionality Reduction for Polynomial and Artificial Neural Network MIMO Adaptive Digital Predistortion

David López-Bueno, *Senior Member, IEEE,* Gabriel Montoro, and Pere L. Gilabert, *Senior Member, IEEE*

***Abstract*—In 5G and beyond radios, the increased bandwidth, the fast-changing waveform scenarios, and the operation of large array multiple-input multiple-output (MIMO) transmitter architectures have challenged both the polynomial and the artificial neural network (ANN) MIMO adaptive digital predistortion (DPD) schemes. This paper proposes training data selection methods and dimensionality reduction techniques that can be combined to enable relevant reductions of the DPD training time and the implementation complexity for MIMO transmitter architectures. In this work, the combination of an efficient uncorrelated equation selection (UES) mechanism together with orthogonal least squares (OLS) is proposed to reduce the training data length and the number of basis functions at every behavioral modeling matrix in the polynomial MIMO DPD scheme. For ANN MIMO DPD architectures, applying UES and principal component analysis (PCA) is proposed to reduce the input dataset length and features, respectively. The UES-OLS and the UES-PCA techniques are experimentally validated for a 2x2 MIMO test setup with strong power amplifier input and output crosstalk.**

***Index Terms*—artificial neural networks, digital predistortion, machine learning, MIMO, power amplifier.**

I. Introduction

THE advent of 5G has brought deploying flexible waveform, numerology and frame design strategies together with increased bandwidth signals that operate in large-array transmitter architectures with RF impairments that cannot be handled properly by classical digital predistortion (DPD) linearizers [1]. Nowadays 5G multi-antenna transmitters may deliver fast-changing waveform configurations, either in terms of modulation, bandwidth occupation or power statistics, that require faster DPD coefficient adaptation. Such transmitters need also to handle complex linearization scenarios involving multiple power amplifier (PA) input and output cross coupling effects and beam dependent PA loading effects, which aggravate the overall PA distortion [2]. Fig. 1 shows the block diagram of one MIMO transmitter architecture for sub-7 GHz and depicts some undesired effects that may appear like cross couplings, I/Q gain and phase imbalances, DC offsets, in-band LO couplings and PA nonlinear distortion [3], [4].

This publication was supported in part by the projects PID2020-113832RB-C22 and PID2020-113832RB-C21 funded by MCIN/AEI/10.13039/501100011033, and in part by the project TSI-063000-2021-121 (MINECO UNICO programme) funded by the European Union-NextGenerationEU through the Spanish Recovery, Transformation and Resilience Plan.

D.López-Bueno is with the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA), Castelldefels, BCN 08860 SPAIN, e-mail: david.lopez@cttc.es

G. Montoro and P. L. Gilabert are with the Department of Signal Theory and Communications, Universitat Politècnica de Catalunya (UPC) - Barcelona Tech, Barcelona, 08034, SPAIN, e-mail: gabriel.montoro@upc.edu, plgilabert@tsc.upc.edu

A. State-of-the-Art

Several MIMO DPD behavioral models accounting for cross couplings have been presented over the last decade [6]–[8], some of them capable to mitigate I/Q modulator imbalances and DC offsets added to the crosstalk effects [9]–[11]. The main concern of all the previous multivariate polynomial-based models is that a multi-input DPD is required in every transmit path and cannot be properly implemented for large-array MIMO transmitters. To overcome the requirement of including a multi-input DPD in every transmit path, a DPD system combining a single linear crosstalk and mismatch model block for the whole array and dual-input DPD models at each Tx path is proposed in [12]. In addition, artificial neural networks (ANN) have been proposed in literature over the last years as an alternative to such higher complexity polynomial models for MIMO DPD [13]–[15]. Such schemes benefit from having only one ANN MIMO DPD block with as many inputs and outputs as transmitter baseband channels and PAs, which reduces the overall count of DPD coefficients for large arrays. The single DPD block approach is also adopted in [16]–[18] for addressing the linearization of hybrid beamforming transmitter architectures at mm-waves.

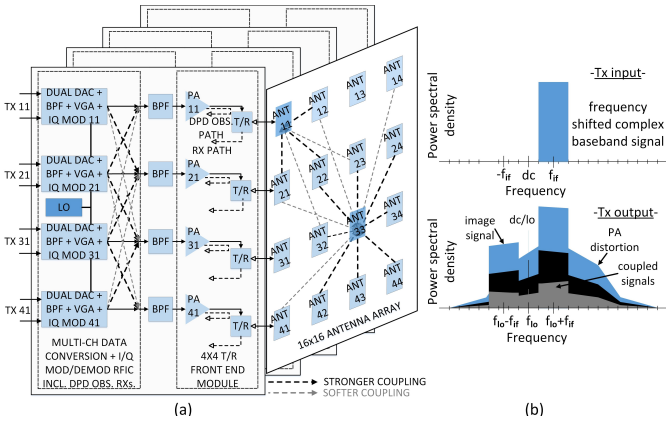

Fig. 1: Multi-antenna circulator-less Tx. architecture (left) and illustration of the spectra of the complex baseband signal and the RF signal at a given PA output with typical RF impairments (right). Source: [5].

The adaptive multi-antenna DPDs can become key building blocks in nowadays radio modems building massive digital MIMO or hybrid beamforming transmitter architectures, but only if designed to be computationally and power efficient. The combination of training data selection methods and behavioral modeling dimensionality reduction techniques is a design strategy that may allow fast-adaptive well-conditioned MIMO DPD systems with optimized DPD processing complexity.

The training data reduction methods reduce the memory requirements, the number of operations and the overall DPD adaptation time, by selecting convenient statistically representative training data samples. Several sample selection methods (SSM) have been reported in literature to reduce the computational complexity of polynomial- or piecewise-based DPD and to uncorrelate the observation errors in adjacent samples to get better performance [19]–[22]. For ANN DPD, consecutive and sparse waveform batch selection mechanisms were proposed in [23] to reduce the training dataset length.

Among the different dimensionality reduction techniques used for polynomial- or piecewise-based DPD, feature selection techniques [24] are employed to reduce the number of DPD basis functions (and thus coefficients) in the real-time forward path. Instead, feature extraction techniques are used in the observation path to ensure well-conditioned estimation and reduce the number of parameters required for DPD identification or adaptation [25]. In ANN-based DPD, feature selection techniques have been used to reduce the input dataset and ANN complexity [26], to tune the center of radial basis function neural networks [27] and to incrementally set the optimal number of hidden layer neurons [28].

Most of DPD complexity reduction works in literature apply to single channel transmit architectures and only a few combine training data selection and basis reduction techniques [23], [29]. Regarding multi-antenna DPD, the authors in [30] use sparse estimation techniques to reduce the basis of MIMO Volterra-based polynomial models for moderate input-output crosstalk conditions. In [31], a piecewise closed-loop DPD including a pruning algorithm for faster adaptation is introduced, while in [32] singular vector decomposition (SVD) is applied for dimensionality reduction of multiuser MIMO arrays.

B. Contribution and Novelty

To the author's best knowledge, the combination of training data selection methods and dimensionality reduction techniques has not been addressed in the literature for multi-antenna transmitter architectures with PA input and output cross couplings provoking strong nonlinear effects. Typically, such techniques have been evaluated in single-antenna systems or multi-antenna systems with negligible cross couplings that allow treating each PA independently. In this work, for the first time several training data selection methods and dimensionality reduction techniques are combined for use in different polynomial and neural network direct learning MIMO DPD architectures to combat strong nonlinear effects and cross couplings.

As proved in Section VI-B, the histogram-based sample or equation selection methods [21], [22] fail to deliver the expected performance in MIMO transmitter architectures when the channel of interest is contaminated by adjacent channels due to PA input and output cross talk. In this work, we evaluate the waveform batch selection methods in [23] whose selection is based in batch performance metrics. More importantly, we propose using the more efficient uncorrelated equation selection (UES) method whose novel application to data selection for least squares (LS) fitting, in our case applied to DPD, outperforms the previous techniques, supports operation under channel or antenna cross couplings, and runs faster.

Regarding the art in dimensionality reduction techniques, we leverage on the novel application of the orthogonal least squares (OLS) and PCA to MIMO DPD. We have deployed the original OLS technique by Chen and Billings [33] for DPD basis selection. This technique has inspired similar approaches validated for single-antenna DPDs. We define how to apply the original OLS technique to reduce the polynomial model basis functions and ANN dataset features in MIMO DPD, knowing that it is able to outperform OMP in DPD basis selection and perform equally well than its later variants [34]. In polynomial-based DPD, PCA is commonly used in the identification subsystem to reduce the number of basis and avoid an ill-conditioned estimation while reducing the complexity of the LS calculation [5]. We propose using PCA to reduce the number of features in large ANN MIMO DPD datasets. PCA is able to extract the hidden structure in high-dimensionality datasets, produce linear combinations of dataset features that help the ANN to reach the desired results for a given structure, and reduce the complexity both in training and inference thanks to reducing the number of weights needed in the ANN first hidden layer. As demonstrated in Section VI-B, PCA and OLS can be used to trade-off between ANN performance and reduced complexity (and faster adaptivity), respectively.

This article formulates and gives full visibility on how the methods for faster MIMO DPD adaptivity are applied to each of the three direct learning architectures, which is a contribution frequently obviated in the literature but necessary to allow reproducibility. The DPD adaptivity or some training aspects in ANN-based digital predistorters are commonly not addressed in the literature, and their limitations in comparison to polynomial DPDs are sometimes omitted. The techniques are combined and experimentally validated in a 2×2 MIMO laboratory test bench for the three MIMO DPD schemes under the presence of strong cross couplings and nonlinear effects.

The remainder of this paper is organized as follows. In Section II the direct learning MIMO DPD schemes for the polynomial and ANN approaches are presented. Section III focuses on the DPD training data length reduction mechanisms and details the new UES method for MIMO architectures with cross couplings. Section IV introduces how to apply OLS and PCA to polynomial and ANN MIMO predistorters. The procedure to combine the proposed training data reduction and dimensionality reduction techniques for MIMO DPD application is detailed in Section V. The test setup employed to experimentally validate the proposed techniques in MIMO DPD architectures is described in Section VI-A. Experimental results showing the benefits in terms of performance, complexity, and training time of combining dimensionality reduction techniques and data

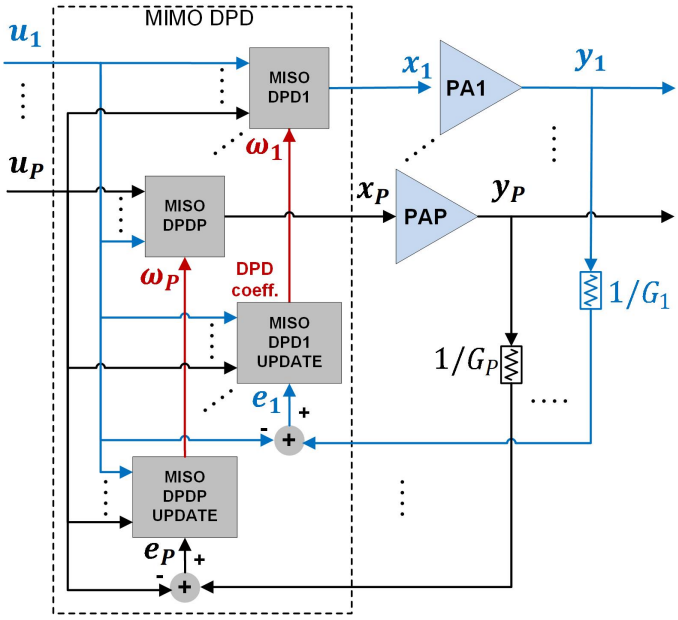

Fig. 2: MIMO DPD direct learning architecture with independent polynomial MISO DPD blocks.

length reduction methods are provided in section VI-B. The conclusion is given in Section VII.

II. Direct Learning Multi-Antenna DPD Schemes

A. Polynomial-Based MIMO DPD Scheme

The block diagram of the polynomial MIMO DPD closed-loop adaptive architecture is shown in Fig. 2. In the forward path of this direct learning MIMO DPD scheme, we have one multiple input single output (MISO) DPD block per PA or antenna, whose input-output relationship can be described as

$$x_i[n] = u_i[n] - d_i[n] \tag{1}$$

where $x_i[n]$ is the predistorted signal at the output of the i^{th} MISO DPD block, with $i = 1, \cdots, P$, P is the number of PAs or antennas (or MIMO baseband signals in a digital MIMO transmitter), $u_i[n]$ is the i^{th} MIMO baseband signal (it is noted that the MIMO input signals are uncorrelated), and $d_i[n]$ is the PA distortion signal to be modeled. The polynomial model will not only account for the i^{th} MIMO baseband signal but all the rest to deal with any PA input and output cross coupling and nonlinear effect. The expansion of the generalized memory polynomial (GMP) model for MIMO scenarios with nonlinear crosstalk (GMPNLC) in [10] is used in this work to model the PA distortion.

It is possible to rewrite (1) in matrix notation as follows

$$\boldsymbol{x_i} = \boldsymbol{u_i} - \boldsymbol{U_i}\boldsymbol{w_i} \tag{2}$$

where $\boldsymbol{x_i} = \left(x_i[0], \cdots, x_i[n], \cdots, x_i[L-1]\right)^T$ and $\boldsymbol{u_i} = \left(u_i[0], \cdots, u_i[n], \cdots, u_i[L-1]\right)^T$, with $n = 0, \cdots, L-1$, are the MISO DPD block output predistorted signal and the i^{th} MIMO channel baseband signal input, respectively, and $\boldsymbol{U_i} = \left(\boldsymbol{\phi_i}[0], \cdots, \boldsymbol{\phi_i}[n], \cdots, \boldsymbol{\phi_i}[L-1]\right)^T$ is the $L \times N_i$ behavioral modeling data matrix, with L being the number of samples and N_i being the number of basis functions or the order of the i^{th} MISO DPD model, and where $\boldsymbol{\phi_i}^T[n] = \left(\varphi_{i1}[n], \cdots, \varphi_{ij}[n], \cdots, \varphi_{iN_i}[n]\right)$ is the vector containing the specific basis functions $\varphi_{ij}[n]$ with $j = 1, \cdots, N_i$ that will be applied to the input data to constitute $\boldsymbol{U_i}$. The vector of DPD coefficients with dimensions $N_i \times 1$ is $\boldsymbol{w_i} = (w_{i1}, \cdots, w_{ij}, \cdots, w_{iN_i})^T$. The total number of coefficients in the polynomial MIMO DPD will be $N_t = \sum_{i=1}^{P} N_i$.

In the MISO DPD update paths, the coefficients can be extracted iteratively finding the LS solution. At the k^{th} iteration the coefficients are obtained as

$$\boldsymbol{w_i}^{k+1} = \boldsymbol{w_i}^{k} + \mu\left(\boldsymbol{U_i}^{\mathbf{H}}\boldsymbol{U_i}\right)^{-1}\boldsymbol{U_i}^{\mathbf{H}}\boldsymbol{e_i} \tag{3}$$

where μ $(0 \leq \mu \leq 1)$ is a weighting factor and $\boldsymbol{e_i} = \left(e_i[0], \cdots, e_i[n], \cdots, e_i[L-1]\right)^T$ is the $L \times 1$ DPD error vector defined as

$$\boldsymbol{e_i} = \frac{\boldsymbol{y_i}}{G_i} - \boldsymbol{u_i} \tag{4}$$

where G_i is the desired linear gain of the i^{th} PA, $\boldsymbol{y_i}$ is the signal at its output, and $\boldsymbol{u_i}$ is the i^{th} MIMO baseband signal.

B. Artificial Neural Network MIMO DPD Schemes

As shown in Fig. 3, the ANN-based MIMO DPD can either be built with P independent MISO DPD blocks, each with P inputs and a single output and producing a coefficient vector $\boldsymbol{c_i}$ (as in the polynomial approach), or with a single MIMO DPD processing block with P inputs and P outputs and a unique coefficients vector $\boldsymbol{c}$.

In the MIMO DPD built with independent ANN MISO DPD blocks, every block is fed with all the baseband channel inputs $\boldsymbol{u_i}$ (in practice, only those with cross-effect interaction should be accounted) and every set of coefficients $\boldsymbol{c_i}$ is calculated to produce the ANN expected output (target signal) $\boldsymbol{x'_i} = \boldsymbol{u_i} - \boldsymbol{e_i}$ that precompensates for the distortion added by the corresponding PA and accounts for linear and nonlinear cross couplings from other MIMO channels. Therefore, following the notation in Fig. 3-top, the P sets of $\boldsymbol{c_i}$ coefficients are calculated independently to minimize every error increment $\boldsymbol{\Delta e_i} = \boldsymbol{e_i} - \hat{\boldsymbol{e}}_{\boldsymbol{i}}$, where $\boldsymbol{e_i}$ is the residual linearization error vector at the i^{th} PA output as defined in (4) and $\hat{\boldsymbol{e}}_{\boldsymbol{i}} = \boldsymbol{x'_i} - \hat{\boldsymbol{x}}'_{\boldsymbol{i}}$ is the ANN estimated residual linearization error vector between the target signal and the estimated signal $\hat{\boldsymbol{x}}'_{\boldsymbol{i}}$.

The MIMO DPD built with the single ANN MIMO DPD block in Fig. 3-bottom jointly generates all the precompensation signals minimizing the error between every $(\boldsymbol{e_i}, \hat{\boldsymbol{e}}_{\boldsymbol{i}})$ pair, and thus $\boldsymbol{c}$ is now calculated to minimize the sum of all the error increments $\boldsymbol{\Delta e_i}$. The topology shown in Fig. 4 is proposed for this architecture. It is an augmented version of the feedforward fully connected MIMO-RVTDNN in [14]. This ANN has real-valued input IQ pairs defined as the real and imaginary part of the input $u_i[n]$, where $I_{\text{in},i}[n] = Re\{u_i[n]\}$ and $Q_{\text{in},i}[n] = Im\{u_i[n]\}$, with $i = 1, \cdots, P$ (one input IQ pair per antenna), and real-valued output IQ pairs with $\hat{I}'_{\text{out},i}[n] = Re\{\hat{x}'_i[n]\}$ and $\hat{Q}'_{\text{out},i}[n] = Im\{\hat{x}'[n]\}$, where $\hat{x}'[n]$ is the ANN estimation of the targeted output $x'[n]$ with $I'_{\text{goal},i}[n] = Re\{x'_i[n]\}$ and $Q'_{\text{goal},i}[n] = Im\{x'_i[n]\}$

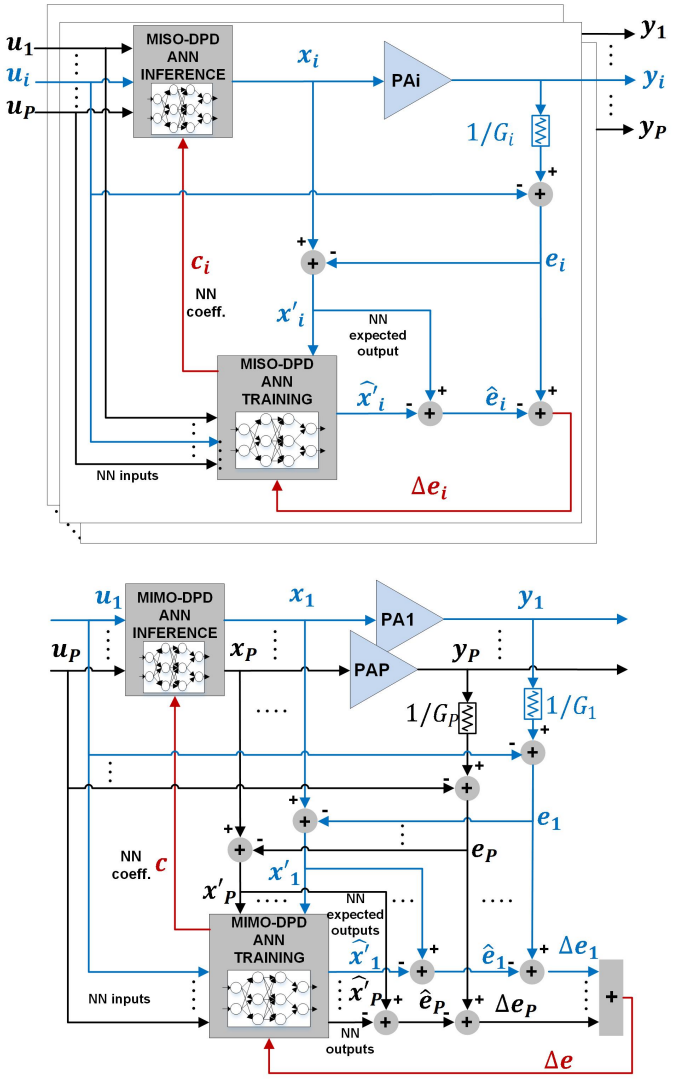

Fig. 3: MIMO DPD direct learning architecture with independent ANN MISO DPD blocks (top) and with a single ANN MIMO DPD block (bottom).

components. Every input IQ pair is augmented by several data functions (DF in Fig. 4) that enrich the ANN basis functions to improve nonlinear modeling. One of the augmentations that favor nonlinear modeling is using DFs that generate envelope dependent terms (i.e., $\left|I_{\text{in},i}+jQ_{\text{in},i}\right|^k$ with $k \in \mathbb{N}$) [15]. Other DFs to be potentially considered provide powers of the IQ data (i.e., $\left|I_{\text{in},i}\right|^k$ and $\left|Q_{\text{in},i}\right|^k$) or angle dependent terms (i.e., $\left(\arctan(Q_{\text{in},i}/I_{\text{in},i})\right)^k$). It is well noted that such augmented products help the ANN to elevate the reachable modeling nonlinear order without having to increase the number of hidden layers using nonlinear activation functions. For proper dynamic nonlinear system identification, tapped-delay lines are added and applied to the input IQ pairs and to the DF outputs (M is the memory depth), delivering the N ANN inputs shown in Fig. 4. Long-term memory effect modeling components (i.e., $s[n] = \frac{1}{K}\sum_{k=0}^{K-1}|I_{ini}[n-k]+jQ_{ini}[n-k]|^2$) can also be introduced by the data functions but these inputs do not need to be replicated and delayed by the tapped-delay system.

All the previous ANN inputs constitute the input dataset features. Considering the input IQ data length L, the input dataset will have $N \times L$ size. The size of the dataset will be impacted by the complexity of the dynamic nonlinear effects to be modeled. As observed in Fig. 4, the input dataset $\boldsymbol{U} = (\boldsymbol{D_1}, \cdots, \boldsymbol{D_i}, \cdots, \boldsymbol{D_P})$ is built with the contributions from every input IQ pair or MIMO baseband channel routed into the MIMO DPD ANN. Considering P antennas, V data functions and M sample memory depth, after augmentation and delaying, every MIMO channel contributes with $(2+V)(M+1)$ dataset variables, totaling $N = P(2+V)(M+1)$ ANN inputs. We can now define $\boldsymbol{D_i} = \left(\boldsymbol{\phi_i}[0], \cdots, \boldsymbol{\phi_i}[n], \cdots, \boldsymbol{\phi_i}[L-1]\right)^T$, where $\boldsymbol{\phi_i}^T[n] = \left(\varphi_1^{\boldsymbol{u_i}}[n], \cdots, \varphi_j^{\boldsymbol{u_i}}[n], \cdots, \varphi_{(2+V)(M+1)}^{\boldsymbol{u_i}}[n]\right)$ can be seen as a vector containing all the specific dataset features (or basis functions) $\varphi_j^{\boldsymbol{u_i}}[n]$ linked to a given MIMO input signal $\boldsymbol{u_i}$, for $i = 1, \cdots, P$ and $j = 1, \cdots, (2+V)(M+1)$.

The proposed MIMO DPD ANN is therefore created with one input layer with N inputs, a first and a second hidden layer with F and S neurons, respectively, and an output layer with R neurons and outputs, with $R = 2P$. In the first and second hidden layer we have the $\vartheta^1(\cdot)$ and $\vartheta^2(\cdot)$ activation functions which are typically nonlinear functions such as the hyperbolic tangent sigmoid (well suited for nonlinear modeling), while a pure linear activation function is used at the output layer. The total number of ANN DPD coefficients in $\boldsymbol{c}$ is divided into $FN + SF + RS$ weights and $F + S + R$ biases.

In the approach with independent ANN MISO DPD blocks, the sum of all the $\boldsymbol{c_i}$ sets totals $P(FN+SF+2S)$ coefficients. While the complexity in terms of coefficients will be higher and the training of all the independent ANNs can take longer without hardware (HW) parallelization, the calculation of coefficients at each MISO DPD block will run faster and the performance will be also better due to dividing the problem into smaller ones (i.e., the coefficients at every independent ANN are tuned to minimize a single channel increment error vector and not the sum of potentially unbalanced different channel increment error vectors).

The MIMO DPD ANN output IQ pairs $\hat{I}'_{\text{out},i}[n]$ and $\hat{Q}'_{\text{out},i}[n]$ are calculated during training in the forward pass. The MIMO-RVTDNN coefficients $\boldsymbol{c}$ (ANN weights and biases) are then calculated with a backpropagation algorithm. Taking as reference the Levenberg-Marquardt (LM) algorithm [35] and the fully integrated MIMO DPD ANN, when going backward the coefficients are calculated by minimizing the mean square error (MSE) cost function that relies on the sum of all the $\boldsymbol{\Delta e_i}$. The forward-backward procedure is iterated until the required linearization performance is achieved or the ANN fails in generalization.

This procedure is equivalently applied P times for the MISO ANN DPD. The $\boldsymbol{U}$ input dataset will be shared between all these independent ANNs but now the single target goals $\boldsymbol{x'_i}$ will be different according to the baseband channel of interest and the effects to be counteracted at each of the corresponding transmit chains. The cost function instead of minimizing the sum of error increments it will now minimize the error increment $\boldsymbol{\Delta e_i}$ for a given channel (see Fig. 3-top).

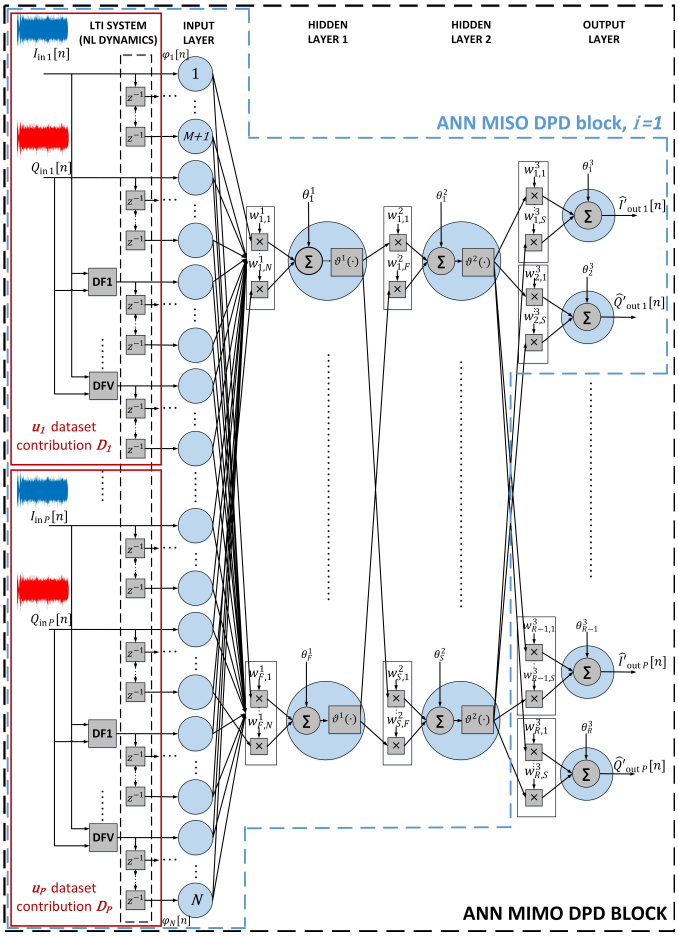

Fig. 4: Two-hidden layer data function augmented fully connected RVTDNN for MIMO DPD application.

III. Training Data Length Reduction

A. Batch and Equation Selection Methods

In [23] two batch selection methods named consecutive batch selection (CBS) and sparse batch selection (SBS) were presented by the authors of this work. These methods were used to reduce the length of the ANN input dataset for nonlinear DPD modeling. These methods are compared with a SSM-inspired multidimensional IQ memory mesh selecting (MeS) method in [22] that can be applied to the equations (or rows) of either polynomial behavioral modeling matrices or ANN input datasets. Compared to MeS, CBS and SBS could also achieve high reduction factors, similar normalized mean square error (NMSE) performance but better adjacent channel power ratio (ACPR) thanks to employing batch selection score metrics linked to the out-of-band performance such as the adjacent channel error power ratio (ACEPR). One advantage of the batch selection methods is the versatility to adapt to different scenarios and properly balance the in-band and out-of-band distortion according to the requirements. For completeness, the application of CBS and SBS to single channel DPD input-output data is summarized and exemplified in Fig. 5.

When these batch and equation selection techniques are applied to ANN MIMO DPD the following modifications can be applied to guarantee suitable time coexistence of the

1. $\boldsymbol{u}$ and $\boldsymbol{y}$ are divided in N_b data batches. N_b is integer multiple of the reduction factor RF_l. Batches to be selected: $N_s = N_b/RF_l$.

2. **CBS:** metric calculation over sliding window with N_s batch width (length of $L_s = L/RF_l$ samples), shifted over $\boldsymbol{u}$ and $\boldsymbol{y}$ in $L_b = L/N_b$ sample steps. **SBS:** metric calculation over every consecutive batch (N_b in total).

3. Tunable batch scoring dep. on: a) **worst NMSE** b) **worst ACEPR** c) **highest $\boldsymbol{u}$ rms PWR**.

4. **CBS:** Single batch selection. **SBS:** Batch indices are sorted descending according to the score, and the selection is done with highest scoring ones. These are sorted ascending.

5. **CBS and SBS**: The batch indices are expanded to include the indices of the samples that contain, and form $\boldsymbol{\beta}$.

Example: $RF_l = 2, N_b = 4, N_s = 4, L = 40 \rightarrow L_s = 20, L_b = 10$

CBS

Batch calc->	1	2	3
NMSE score	3	2	1
ACEPR score	1	3	2
PWR_u score	2	3	1
Batch score	**6**	**8**	**4**

max_score → index=[2]

$\boldsymbol{\beta}$=((index-1)* L_b + 1: (index−1)∗L_b + L_s)= =(11:30)

SBS

Batch calc->	1	2	3	4
NMSE score	3	2	1	4
ACEPR score	1	4	2	3
PWR_u score	1	3	2	4
Batch score	**5**	**9**	**5**	**11**

[4,2,1,3] -> [4,2] → indices=[2,4]

$\boldsymbol{\beta}_k$=((indices(k)-1)* L_b + 1: indices(k)∗L_b)

for k=1,2 → $\boldsymbol{\beta}$=[$\boldsymbol{\beta}_1$ $\boldsymbol{\beta}_2$]=[(11:20) (31:40)]

Fig. 5: CBS and SBS batch selection procedure [23].

selected training data samples between every MIMO channel dataset contribution $\boldsymbol{D_i}$ in $\boldsymbol{U}$:

- For CBS and SBS, when the scores are calculated for every $(\boldsymbol{u_i}, \boldsymbol{y_i})$ pair, these are then summed, and a unique global batch index selection vector $\boldsymbol{\beta}$ is obtained for all the MIMO channel input dataset contributions. Given a length reduction factor RF_l, $\boldsymbol{\beta}$ will store the relevant L/RF_l training data indices that will be applied to the input-output data. This strategy can also be applied to the polynomial MIMO DPD case allowing for higher reduction factors. Under dominant strong cross couplings, only the mean power of the $\boldsymbol{u_i}$ signal batches is needed to select the right indices, which reduces the number of operations.
- With MeS the desired reduction factor RF_l is expanded by 50% and more equation indices than those initially required are selected. Once these indices are selected for all channels independently, only those appearing at all channels are taken and stored in the vector $\boldsymbol{\Theta}$. Finally, the number of coincident indices is reduced to provide the desired L/RF_l equation selection indices. This strategy is not required for the polynomial MIMO DPD.

These MIMO adaptations applied to the CBS, SBS and MeS techniques are also deployed in Section V. From now onward, we will assume that the same length reduction factor is applied to any MISO polynomial or MISO ANN block, i.e. $RF_{li} = RF_l$ for $i = 1, ..., P$). As shown in the experimental results in Section VI, the traditional histogram-based methods can be underperforming for MIMO DPD application since the baseband sample selection made according to the histogram distribution of the baseband signal is no longer representative of what happens at PA level due to having uncorrelated couplings at the MIMO transmitter both before and after the PA. SBS and CBS, however, can be more suitable to select appropriate waveform segments thanks to tuning the performance scoring method to reflect, for instance, the batches where the waveforms are more impacted by the crosstalk effects.

Algorithm 1 Uncorrelated equations selection (UES) method

procedure UES $(\boldsymbol{U}, CF)$
  initialization:
    $\boldsymbol{r} \leftarrow \boldsymbol{U}(1,:)$
    $\boldsymbol{\Theta}^{(0)} \leftarrow \{1\}$
  for $m = 2$ **to** L **do**
    if $|\boldsymbol{U}(m,:)\ \boldsymbol{r}^H| < CF$ **then**
      $\boldsymbol{r} \leftarrow \boldsymbol{U}(m,:)$
      $i^{(m)} \leftarrow m$
      $\boldsymbol{\Theta}^{(m)} \leftarrow \boldsymbol{\Theta}^{(m-1)} \cup i^{(m)}$
    end if
  end for
  return $\boldsymbol{\Theta}$
  return $\boldsymbol{U}_{\{\boldsymbol{\Theta}\}}$
end procedure

B. Uncorrelated Equations Selection Method

A simple and computationally efficient equation selection method is proposed in this work for application to both the $\boldsymbol{U_i}$ polynomial behavioral modeling matrices and the $\boldsymbol{U}$ ANN input dataset. The procedure is detailed in Algorithm 1. It starts storing the first equation (i.e., matrix row) in $\boldsymbol{U}$ as the original master equation $\boldsymbol{r}$, and index one is stored in the equation selection index vector $\boldsymbol{\Theta}$. The scalar product between the master equation and the forthcoming equations is then applied. When one of such products is below the correlation factor (CF) threshold, the new master equation is stored together with its index. The process is repeated until all the equations are evaluated, the equation selection index vector $\boldsymbol{\Theta}$ is completed and it is then applied to $\boldsymbol{U}$ to reduce the number of equation or rows. The CF value to be set depends on the desired length reduction factor RF_l, i.e., the higher RF_l is, the smaller the CF value will be.

This method shows a good trade-off between implementation complexity and reduction performance. Such a procedure not only reduces the number of correlated equations, and thus provides better system conditioning but it is also able to identify samples that are impacted by cross couplings. This is because the correlation between adjacent OFDM waveform samples for a given MIMO channel would be higher without crosstalk effects, and thus by sequentially selecting uncorrelated equations those highly impacted by crosstalk effects are stored. Fig. 6 shows an example on how the equation and batch selection methods look like when applied to one input baseband signal with $RF_l = 40$. While MeS fills in most of the histogram bins with the samples or equations appearing initially in the training data, UES takes them in a more uniform fashion.

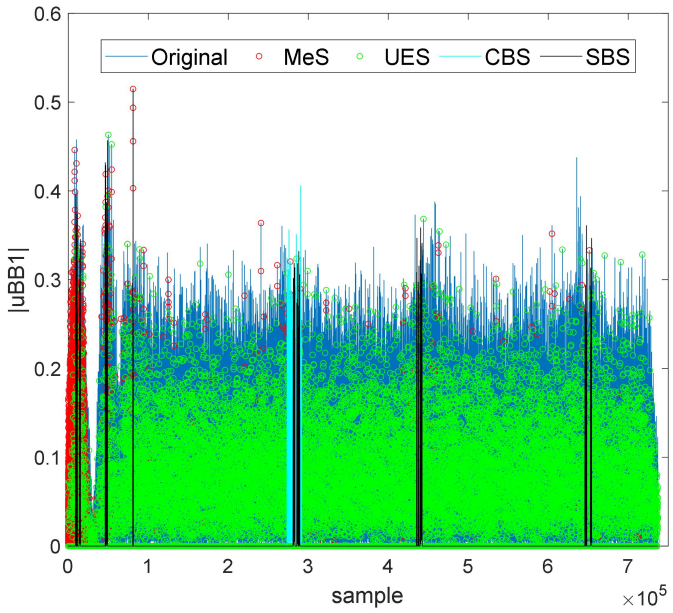

Fig. 6: Training data length reduction methods application.

IV. DPD Model Order Reduction

A. Orthogonal Least Squares

Several feature selection techniques have been proposed for DPD linearization to keep the most representative basis functions (or regressors) and thus reduce the modeling dimensionality [24]. From the matching pursuit family, the orthogonal matching pursuit (OMP) [36] and the doubly OMP (DOMP) [34] have proved to be suitable approaches to trade-off linearity performance and model order reduction. In this work, the original orthogonal least squares (OLS) technique shown in Algorithm 2 was considered because it outperforms OMP and we are facing a challenging nonlinear scenario.

Algorithm 2 is defined generically. The cumulative modeling error $\boldsymbol{e}$ relies on a target signal $\boldsymbol{t}$. This target signal will be the PA output $\boldsymbol{y_i}$ when using the polynomial MIMO DPD scheme in Fig. 2 and $\boldsymbol{x_i'}$ when using ANN-based schemes like those in Fig. 3. As in OMP, in OLS it is possible to define a support vector ($\boldsymbol{\Upsilon}$) containing the indices of the best basis functions of a generic polynomial behavioral modeling matrix or ANN input dataset $\boldsymbol{U}$. At every iteration of the Algorithm 2 search, $\boldsymbol{\Upsilon}$ will be fed with the indices corresponding to the basis functions that better contribute to minimize the residual modeling error, which are sorted according to their relevance. For a given dimensionality reduction factor RF_d, the algorithm iterates N/RF_d times to fill $\boldsymbol{\Upsilon}$ with the required number of indices. This support set will be finally applied to the $\boldsymbol{U}$ matrix to obtain a reduced version having only the selected basis. Unlike OMP, OLS adds deflation to the chosen columns so when a basis function is chosen, those still to be selected are orthogonal with respect to the chosen ones. Gram-Schmidt is performed by first obtaining a projection vector $\boldsymbol{\rho}$ of the selected regressor into each one of previously orthogonalized regressors in $\boldsymbol{Z}$ (step 11). This projection is then used to decorrelate the basis functions with the selected regressor as per step 12 in Algorithm 2, where $\otimes$ is the Kronecker product operator. The selection step used in OLS is different from that used in OMP, since it accounts for the minimum residual error after orthogonalization.

The specific usage of OLS both for the polynomial and the ANN-based MIMO DPD is detailed as follows. For polynomial MIMO DPD schemes, OLS will be applied at every MISO DPD block as $\boldsymbol{U_i^{OLS}}$ = OLS($\boldsymbol{y_i}, \boldsymbol{U_i}, RF_d$) with

Algorithm 2 Orthogonal Least Squares

procedure OLS $(\boldsymbol{t}, \boldsymbol{U}, RF_d)$
  initialization:
  $\boldsymbol{e}^{(0)} \leftarrow \boldsymbol{t}_{L\times 1}$
  $\boldsymbol{\Upsilon}^{(0)} \leftarrow \{\}$
  $\boldsymbol{Z}^{(0)} \leftarrow \boldsymbol{U}_{L\times N}$
  for $m = 1$ **to** N/RF_d **do**
    $\boldsymbol{Z}_{\{i\}}^{(m-1)} \underset{i\notin\boldsymbol{\Upsilon}^{(m-1)}}{\leftarrow} \frac{\boldsymbol{Z}_{\{i\}}^{(m-1)}}{\|\boldsymbol{Z}_{\{i\}}^{(m-1)}\|_2}$
    $i^{(m)} \leftarrow \underset{i\notin\boldsymbol{\Upsilon}^{(m-1)}}{\operatorname{argmin}} \min_{\boldsymbol{w_i}} \|\boldsymbol{e}^{(m-1)} - \boldsymbol{Z}_{\{i\}}^{(m-1)}\boldsymbol{w_i}\|_2^2 \approx$
    $\approx \underset{i\notin\boldsymbol{\Upsilon}^{(m-1)}}{\operatorname{argmax}} |\boldsymbol{Z}_{\{i\}}^{(m-1)H}\boldsymbol{e}^{(m-1)}|$
    $\boldsymbol{\Upsilon}^{(m)} \leftarrow \boldsymbol{\Upsilon}^{(m-1)} \cup i^{(m)}$
    $\boldsymbol{e}^{(m)} \leftarrow \boldsymbol{e}^{(m-1)} - (\boldsymbol{Z}_{\{i^{(m)}\}}^{(m-1)H}\boldsymbol{e}^{(m-1)})\boldsymbol{Z}_{\{i^{(m)}\}}^{(m-1)}$
    $\boldsymbol{\rho}^{(m)} \leftarrow \boldsymbol{Z}_{\{i^{(m)}\}}^{(m-1)H}\boldsymbol{Z}^{(m-1)}$
    $\boldsymbol{Z}^{(m)} \leftarrow \boldsymbol{Z}^{(m-1)} - \boldsymbol{\rho}^{(m)} \otimes \boldsymbol{Z}_{\{i^{(m)}\}}^{(m-1)}$
  end for
  return $\boldsymbol{\Upsilon}$
  return $\boldsymbol{U}_{\{\boldsymbol{\Upsilon}\}}$
end procedure

$i = 1, \cdots, P$ (being P the total number of PAs, antennas or MIMO channels). For the ANN approach built with independent MISO DPD blocks, OLS is applied at every MISO DPD block as $\boldsymbol{U_i^{OLS}}$ = OLS($\boldsymbol{x'_i}, \boldsymbol{U}, RF_d$), where $\boldsymbol{U} = (\boldsymbol{D_1}, \cdots, \boldsymbol{D_i}, \cdots, \boldsymbol{D_P})$ is the common MISO ANN's input dataset (i.e., for a small number of antennas all the MIMO baseband channels are routed into any MISO DPD ANN and therefore there is a single $\boldsymbol{U}$ dataset shared between all the MISO DPD ANNs before applying reduction. It is noted that every MISO DPD ANN would have different input datasets otherwise, depending on the physical coupling between channels at every PA or antenna. Note also that, for the sake of simplicity, we have assumed that the same dimensionality reduction factor is applied to any polynomial or ANN MISO DPD block, i.e., $RF_{di} = RF_d$ for $i = 1, ..., P$. In the single-ANN MIMO DPD scheme, OLS is applied independently to every MIMO baseband channel input dataset contribution as $\boldsymbol{U^{OLS}} = \big(OLS(\boldsymbol{x'_1}, \boldsymbol{D_1}, RF_d), \cdots, OLS(\boldsymbol{x'_P}, \boldsymbol{D_P}, RF_d)\big)$, since we have a single dataset $\boldsymbol{U}$ but we need to handle simultaneously the P different target signals $\boldsymbol{x'_i}$. Finally, it is also important to remark that OLS can also be used for basis function or feature preselection to speed-up the UES method, as shown in Section V, thanks to reducing the number of elements per equation and so the number of operations.

B. Principal Components Analysis

This feature extraction technique is suitable for converting an original set of eventually correlated basis functions into a new uncorrelated orthogonal basis set called principal components. For DPD purposes, PCA is typically used to reduce the number of parameters to be estimated in the feedback path and avoid ill-conditioning or over-fitting problems [37]. In polynomial-based MIMO DPD architectures, a new transformed matrix for each MISO DPD block, $\hat{\boldsymbol{U}}_{\boldsymbol{i}}$ (with dimensions $L \times M_i$) is defined as,

$$\hat{\boldsymbol{U}}_{\boldsymbol{i}} = \boldsymbol{U_i}\boldsymbol{V_i} \tag{5}$$

which corresponds to the eigenvectors of the matrix $\boldsymbol{U_i}{\boldsymbol{U_i}}^H$. The $N_i \times M_i$ transformation matrix $\boldsymbol{V_i} = (\boldsymbol{v_{i1}}, \cdots, \boldsymbol{v_{iM_i}})$ is composed of the eigenvectors of the covariance matrix ${\boldsymbol{U_i}}^H\boldsymbol{U_i}$, where N_i is the number of original basis (or DPD coefficients) of the i^{th} MISO DPD block ($i = 1, \cdots, P$), and M_i is the number of transformed basis after reduction (i.e, $M_i < N_i$). When the transformed matrix $\hat{\boldsymbol{U}}_{\boldsymbol{i}}$ is applied to (3), a new reduced set of transformed coefficients $\hat{\boldsymbol{w}}_{\boldsymbol{i}}$ with dimension $M_i \times 1$ is obtained. The original $N_i \times 1$ vector of coefficients is obtained through the transformation matrix as

$$\boldsymbol{w_i} = \boldsymbol{V_i}\hat{\boldsymbol{w}}_{\boldsymbol{i}}. \tag{6}$$

Both the original number of basis N_i and the pruning or reduction factor RF_{di}=N_i/M_i could be different at each MISO DPD block if every PA showed different distortion characteristics. However, when the operating conditions of the PAs are similar the same parameters and reduction factor RF_d can be applied to all the MIMO branches. The transformation matrices $\boldsymbol{V_i}$ are calculated only once and will be valid irrespective of the changing transmitted waveforms when the PA operating conditions do not change very significantly over time. This process can be done a priori offline (i.e., factory profiling) to avoid HW resource utilization and extending the DPD adaptation time, for a given set of representative PA operating conditions. Thanks to the orthogonality of the resulting transformed matrices, the polynomial MISO DPD coefficients' extraction can be carried out with simple dot products (i.e., avoiding matrix inversions), and one-by-one increasing the number of independent components until the desired performance is reached, as shown in [38].

Unlike in polynomial-based approaches, where PCA is not used to directly reduce the number of the original basis functions in the DPD function, in the case of the ANN DPD, PCA does contribute to reduce the size of the ANN. When applied to the input dataset, the ANN will directly operate with the feature-reduced transformed dataset both during training and inference (or feedback and forward paths, respectively). By reducing the input dataset features and so the MIMO ANN inputs and the total amount of data to be processed by the ANN, the number of weights in the first hidden layer FN will be decreased and the overall DPD training or update time will be also reduced. Given the ANN architecture in Fig. 4, PCA can be applied to every MIMO baseband channel input dataset contribution $\boldsymbol{D_i}$. Therefore, considering a MIMO or MISO ANN generic input dataset $\boldsymbol{U} = (\boldsymbol{D_1}, \cdots, \boldsymbol{D_i}, \cdots, \boldsymbol{D_P})$ with $i = 1, \cdots, P$, each transformation matrix $\boldsymbol{V_i}$ can be calculated and then be applied to obtain every $\hat{\boldsymbol{D}}_{\boldsymbol{i}}$ transformed matrix, by using the same procedure as for the polynomial MISO-DPD architecture. After applying PCA as $\boldsymbol{U^{PCA}} = \big(\text{PCA}(\boldsymbol{D_1}, RF_d), \cdots, \text{PCA}(\boldsymbol{D_P}, RF_d)\big)$, the resulting transformed input dataset will be $\hat{\boldsymbol{U}} = \boldsymbol{U^{PCA}} = \big(\hat{\boldsymbol{D}}_{\boldsymbol{1}}, \cdots, \hat{\boldsymbol{D}}_{\boldsymbol{P}}\big)$, both for the MISO (i.e., when all channels are routed into every MISO-ANN DPD block) and the MIMO ANN architectures. Applying PCA to every MIMO channel

input dataset contribution works better than applying PCA to the overall multi-channel dataset since it helps to better preserve the most relevant variables of features provided by every channel. One advantage of PCA over OLS when applied to the input dataset, is that PCA is used as an unsupervised learning technique that helps extracting the hidden structure from the high dimensional dataset. PCA is able to produce linear combinations of the dataset features which helps the ANN reaching the desired results for a given structure.

V. MIMO Scenario Combined Application Procedure

This section details how the training data length reduction methods and the dimensionality reduction techniques are sequentially combined. To do that, Fig. 7 shows the application procedure for the MIMO DPD based on independent polynomial MISO DPD blocks (abbreviated as MISO POLY), while Fig. 8 and Fig. 9 show the procedure for the MIMO DPD with a single ANN MIMO DPD block (MIMO ANN), and the variant with independent ANN MISO DPD blocks (MISO ANN), respectively. The flowcharts show how the PA's input-output data is employed for each architecture defined in Section II, and how the polynomial behavioral modeling matrices and the ANN input datasets (also defined in Section II) are processed according to the length reduction and dimensionality reduction methods deployed in Section III and Section IV, respectively. To reduce the complexity and processing time of the dimensionality reduction techniques the training data length reduction is first applied. In the notation being used, when a length reduction technique is applied to a variable the lr superscript is added to the variable name, while the lr_dr superscript is used when the dimensionality reduction techniques are applied to the length reduced variables.

In a first stage, if data batch selection is chosen then either SBS or CBS is applied to every $(\boldsymbol{u_i}, \boldsymbol{y_i})$ pair. The indices selection in $\boldsymbol{\beta}$ is then applied to such variables, and also to $\boldsymbol{x'_i}$ for ANN training, before building the reduced behavioral modeling and dataset matrices $\boldsymbol{U_i^{lr}}$ and $\boldsymbol{U^{lr}}$, respectively. These contain basis functions that will be a length reduced version of the original ones with L/RF_l samples each. If, otherwise, an equation selection method is chosen, the full behavioral modeling and dataset matrices, $\boldsymbol{U_i}$ and $\boldsymbol{U}$, are first created. With MeS, the method is applied P times to the $\boldsymbol{U_i}$ MISO DPD polynomial blocks or directly to the channel dataset contributions $\boldsymbol{D_i}$ in the ANNs. In either case, P equation selection index vectors $\boldsymbol{\Theta_i}$ are generated. The polynomial MIMO DPD will employ each one at its corresponding MISO-DPD block, while for the ANN dataset these index vectors will be merged into a single one by following the procedure described in Section III-A. As seen in Section III-B, UES is applied to every MISO DPD behavioral modeling matrix $\boldsymbol{U_i}$, and just once to the $\boldsymbol{U}$ ANN dataset. An exception to the latter case is in the MISO ANN architecture when, in order to speed up UES, OLS is applied in advance to reduce the number of columns. Since OLS will be applied to $\boldsymbol{U}$ at every MISO ANN DPD block considering different target signals $\boldsymbol{x'_i}$, we can have different OLS $\boldsymbol{\Upsilon_i}$ support sets

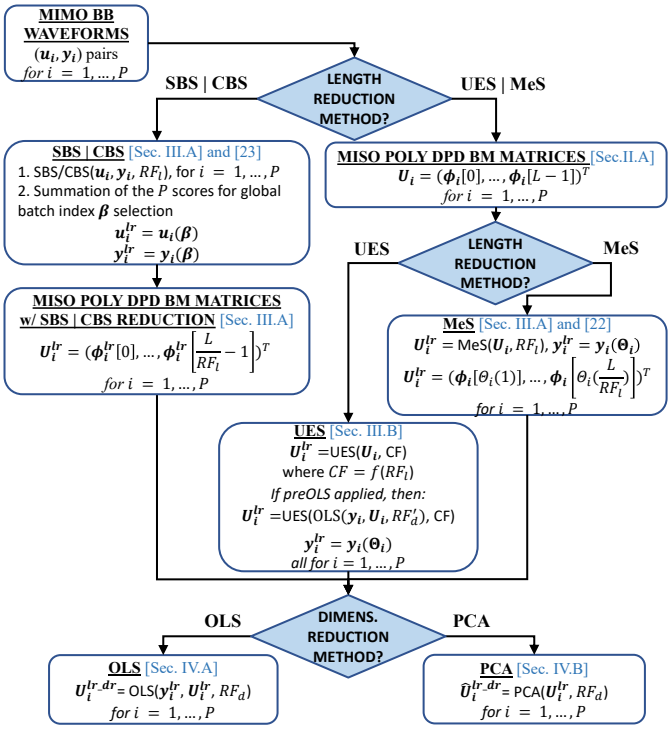

Fig. 7: MISO POLY combined application flowchart.

leading to different equation selections vectors $\boldsymbol{\Theta_i}$ at each of the P MISO ANN blocks. The columns selection can just be used to calculate the equation selection and not be propagated to the next stages. The reduction factor $(RF_d^{'})$ at the preOLS step before UES may be different to the one used for effective dimensionality reduction at the later OLS stage (RF_d).

In a second stage the modeling basis or dataset variable reduction techniques OLS and PCA are applied as described in Section IV-A and Section IV-B, respectively. The final reduced polynomial or ANN input modeling data is then obtained. As previously mentioned, both the preOLS and the OLS column indices may be calculated for every MIMO channel in advance for a given set of operating conditions to avoid increasing the MIMO DPD training/update times. Same applies to the PCA transformation matrices. Finally, it is noted that when a factor RF_d is applied in OLS or PCA to reduce all the polynomial MIMO DPD basis functions, the total number of coefficients is also reduced by this factor. In ANNs, however, this is no longer valid since the OLS/PCA reduction is applied to the input dataset features and thus only the number of weights in the first hidden layer are reduced proportionally.

VI. Experimental Results

A. Test Setup

The proposed schemes have been benchmarked and validated with the MATLAB controlled HW test bench shown in Fig. 10. To launch in parallel multiple DPD experiments a MATLAB client-server waveform upload/download architecture has been enabled. The MATLAB server communicates both with the remote clients running the DPD algorithms and the laboratory HW composed of FPGA boards

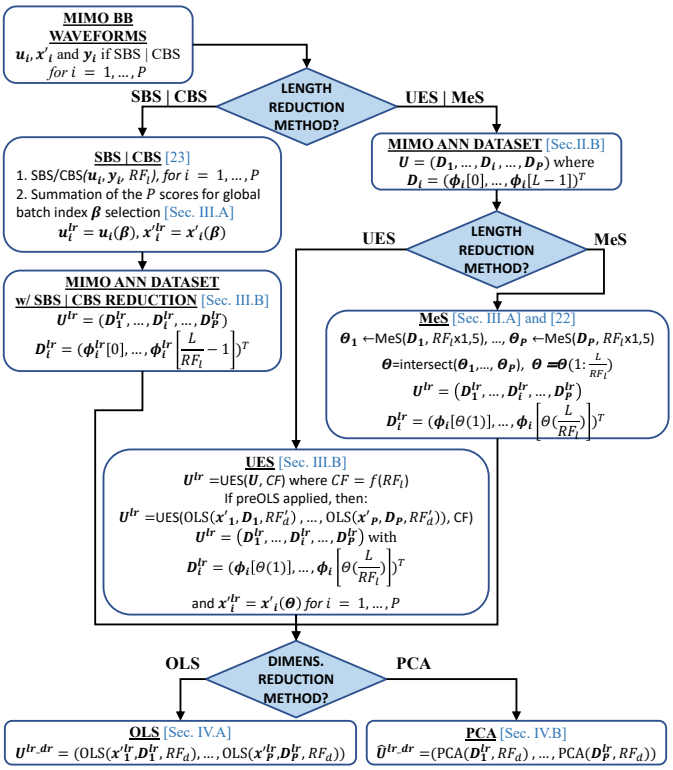

Fig. 8: MIMO ANN combined application flowchart.

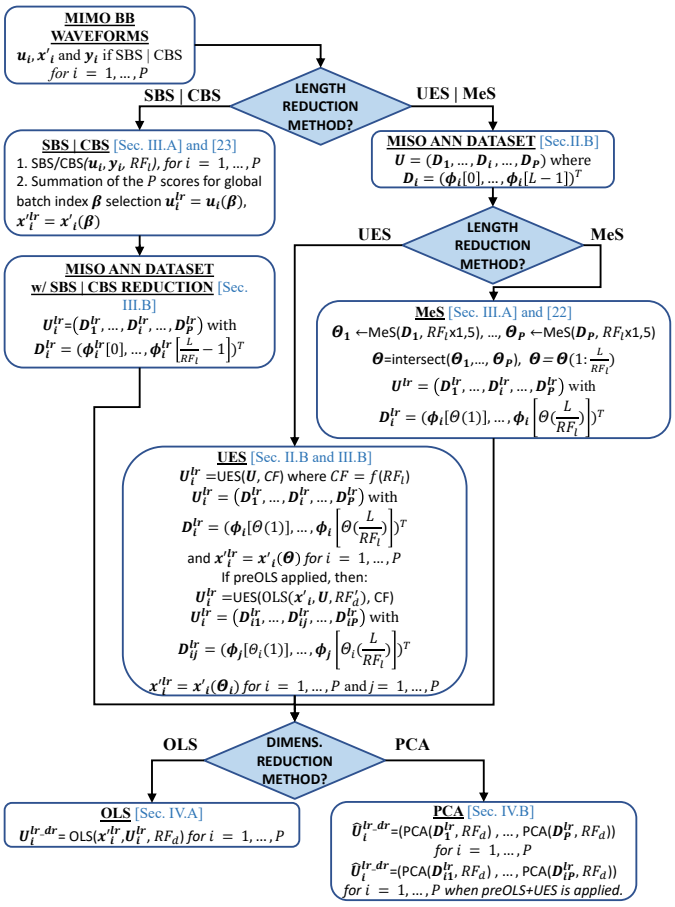

Fig. 9: MISO ANN combined application flowchart.

for waveform playback (TI TSW14J56EVM) and recording (TI TSW14J57EVM), data converters (TI DAC38RF82EVM and TI ADC12DJ5200EVM), two ADI ADL5605 PA drivers, two GaN HEMT class J PAs based on the Cree CGH35030F transistor, and a tunable bidirectional passive network to provoke controlled couplings between the PA outputs and potential loading effects, (by manually setting the attenuation step of the mechanical variable attenuator shown in Fig. 10). The PA input cross talk effects are digitally introduced into the predistorted baseband signals, in the MATLAB Laboratory server before transmission through the test setup, according to the PA input cross talk level set by the remote user.

The PAs are operated at 875 MHz RF frequency and about 27 dBm mean output power with 2x2 MIMO 80 MHz bandwidth carrier-aggregated fast convolution filter bank multicarrier (FC-FBMC) signals featuring 13-14 dB PAPR. The waveform length is 737280 samples, after oversampling 2 ms signals by a factor of 3 to accommodate for the DPD bandwidth expansion (368.64 MHz baseband sample rate). Each of the two FC-FBMC signals generated are uncorrelated and feature changing PAPR values at every training iteration.

The two discrete PAs need some initial adjustment to reach similar operation point due to transistor pinch-off voltage dispersion or having different parasitic effects in the two prototyped boards. The differences in linearity performance imply that there will be some common DPD modeling basis valid for the two models but to reach the maximum performance there will be basis that are needed for one of the PAs but not necessarily the other. Since the general approach is to apply the same overall DPD model to the two PAs, to reduce the DPD complexity and ensure well-conditioned identification the dimensionality reduction techniques are highly convenient.

B. Data Length and Dimensionality Reduction Measurements

The proposed techniques have been experimentally benchmarked and validated in Table I and Table II. The parameters of the three MIMO DPD architectures evaluated are as follows:

1) **MISO POLY**. The MISO DPD block for every PA is based on the GMPNLC model with nonlinear degree 7 and up to 10-tap memory depth. Nonlinear degree 9 and up to 12-tap memory depth was also evaluated but the improvement in performance was not relevant despite nearly doubling the number of coefficients.
2) **MISO ANN**. The MISO DPD blocks are based on two-hidden layer ANNs, with 20 neurons per hidden layer and hyperbolic tangent sigmoid activation functions. The following data augmentation functions are considered for every MIMO channel input dataset (I_{ini}, Q_{ini}) contribution: $|I_{ini} + jQ_{ini}|^k$ with $k = 1, \cdots, 6$, $|I_{ini}|^k$ and $|Q_{ini}|^k$ for $k = 3$, and a long-term component considering a sliding window of $K = 1000$ samples. Except for the long-term component, 12-tap consecutive memory delays are also applied to all these contributions to enable dynamic nonlinear modeling.
3) **MIMO ANN**. The single MIMO DPD block is based on a two-hidden layer ANN, with 20 neurons per hidden

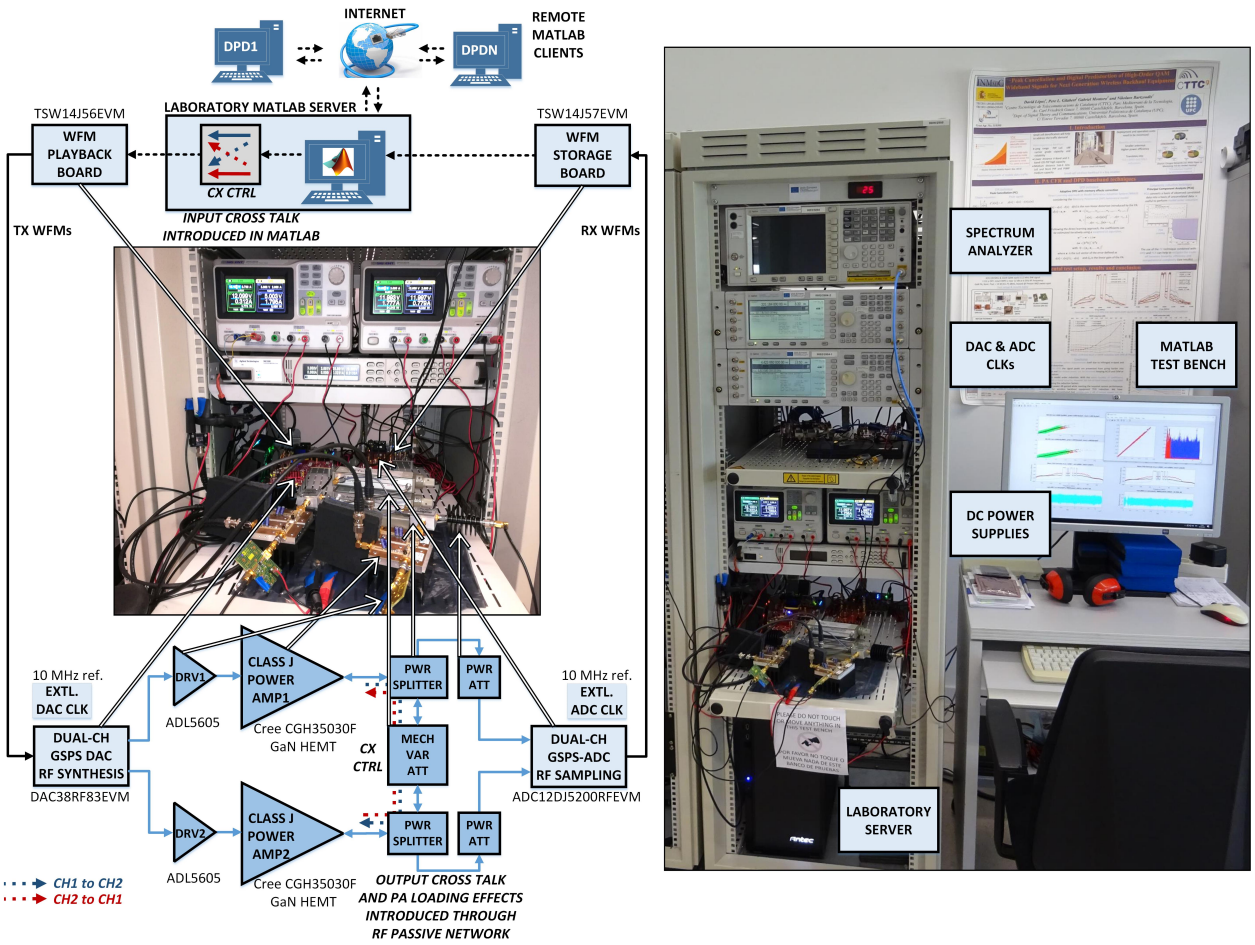

Fig. 10: Remotely accessible multi-antenna digital linearization test setup block diagram (left) and demonstration test bench picture (right). TX: transmitted; RX: received; WFMs: waveforms; CH: channel; EXTL: external; CLK: clock; DAC: digital-to-analog converter; DRV: driver; AMP: amplifier; PWR: power; CX CTRL: cross talk control; MECH VAR ATT: mechanical variable attenuator; ADC: analog-to-digital converter; DC: direct current.

layer and hyperbolic tangent sigmoid activation functions. The following data augmentation functions are added to every MIMO channel input dataset (I_{ini}, Q_{ini}) contribution: $|I_{ini} + jQ_{ini}|^k$ with $k = 1, \cdots, 6$, $|I_{ini}|^k$ and $|Q_{ini}|^k$ for $k = 7, \cdots, 9$ and a long term component considering a sliding window of $K = 1000$ samples. Again, 12-tap consecutive memory delays are applied.

These three MIMO DPD architectures and the researched techniques have been evaluated considering -15 dB PA input and -15 dB PA output cross couplings. Fig. 11 shows the degradation in the NMSE and ACPR figures with increasing equal input and output cross coupling values, when no MIMO DPD is applied. At -15 dB cross coupling level additional nonlinear effects appear due to changes in the input signal power statistics and the class-J PA loading conditions. At such coupling level, and considering the highest PAPR MIMO baseband signals, the PA input signal peaks are compressed about 3 dB at the PA output. No crest factor reduction techniques were applied to the MIMO waveforms. Such challenging scenario requires handling complex MIMO DPD models.

Table I and Table II show complexity, performance and

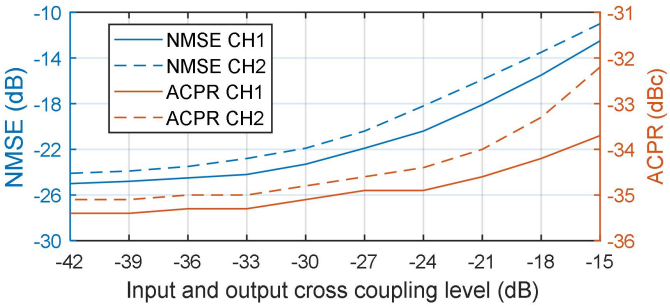

Fig. 11: CH1 and CH2 NMSE and ACPR versus input and output cross coupling level.

timing results (i.e., overall number of MIMO DPD coefficients, NMSE and worst upper/lower ACPR, and time measured with MATLAB's tic toc functions) for a representative set of both dimensionality reduction test cases and the combination of data length and dimensionality reduction test cases, respectively. These techniques are evaluated with the MIMO 2x2 test setup detailed in Fig. 10 for the three MIMO DPD architectures. A number of test case labels are included in the

TABLE I: Dimensionality Reduction Benchmark for MIMO DPD.

	CASE. METHOD	RED. FACT.	DPD COEFF.	NMSE CH2 [dB]	NMSE CH1 [dB]	ACPR CH2 [dBc]	ACPR CH1 [dBc]	TOTAL DPD UPD. TIME [s]		
MISO POLY GMPNLC	C1. GMP 1-D	N/A	644	-11.3	-12.5	-36.7	-38.7	615		
	C2. NONE	**N/A**	**1440**	**-36.3**	**-37.4**	**-46.1**	**-46.3**	**2862**		
	C3. OLS	**4**	**360**	**-36.4**	**-37.4**	**-45.4**	**-45.8**	**232**		
	C4. OLS	**8**	**180**	**-36.5**	**-37.5**	**-45**	**-45.4**	**74**		
	C5. PCA	8	180	-35.5	-36	-42.4	-43	72	**1^{ST} ITER. DPD TT [s]**	**REMAINING ITER. TT [s]**
MISO ANNs	C6. OLS	6	2724	-36.3	-38.6	-46.9	-47.5	2.81E+04	1.42E+04	1.40E+04
	C7. OLS	**16**	**1604**	**-36.3**	**-36.8**	**-47.8**	**-48**	**1.75E+04**	**1.08E+04**	**6.68E+03**
	C8. PCA	**4**	**3604**	**-36.8**	**-38.8**	**-48**	**-48.7**	**3.84E+04**	**2.45E+04**	**1.38E+04**
MIMO ANN	C9. OLS	6	1764	-35.1	-37.4	-44	-45.7	4.01E+04	3.03E+04	9.82E+03
	C10. OLS	**16**	**964**	**-36.5**	**-37.6**	**-45.3**	**-46**	**1.90E+04**	**1.22E+04**	**6.83E+03**
	C11. PCA	**6**	**1764**	**-36.4**	**-38.3**	**-47.2**	**-47.7**	**5.21E+04**	**2.64E+04**	**2.58E+04**

tables and in the text to facilitate the tracking of the results.

Regarding Table I, when the best classical GMP DPD linearization configuration is applied independently to every MIMO channel (C1), since no crosstalk is accounted in the modelling, the resulting NMSE figures are equivalent to those found in Fig. 11 when DPD is not applied (i.e., the crosstalk is a dominant inband effect which is not corrected, only the ACPR is enhanced by about 3-4 dB at C1). The best GMPNLC configuration (with reasonable number of coefficients) is able to overcome the previous limitations but considering the demanding test setup scenario the NMSE values reach near -37 dB NMSE and -46.2 dB ACPR on average (C2). When using the polynomial MIMO DPD scheme, OLS is the most effective technique to reduce the complexity and the total DPD update time while preserving performance. When a coefficient reduction factor of 8 is applied, OLS features about 1.5 dB and 2.5 dB better NMSE and ACPR, respectively, than PCA (C4 vs. C5). The DPD update time is reduced by a factor circa 40 when compared to the nominal case (C4 vs. C1).

When comparing the performance of the polynomial and the ANN linearization schemes, the MISO ANN is the better performing. The results for the nominal ANN-based MIMO DPD schemes without dimensionality reduction are not provided due to the huge amount of time required to obtain them. When considering the MISO ANN scheme after applying PCA dataset reduction by 4, the NMSE is improved by 0.5-1.5 dB but, more importantly, the ACPR is also improved by 2-2.5 dB, in comparison with the GMPNLC MISO POLY scheme (C4 vs. C1). Table I also shows that both for the MISO and MIMO ANNs, for lower basis reduction factors PCA delivers the best attainable performance, but OLS is a better choice for higher reduction factors (C7 vs. C8).

When comparing the number of coefficients, for small arrays the MISO POLY scheme may feature much less coefficients than the ANN schemes (C4 vs. C10) but, however, when the best performance is pursued, they may feature similar number of coefficients. For example, the MISO ANN scheme with OLS reduction by 16 outperforms GMPNLC in terms of spectral contention as seen in Table I (C2 vs. C7) and the linearized spectra plots in Fig. 12. The AM-AM and the AM-PM characteristics for this case are also shown in Fig. 13.

The main drawback of the ANN schemes is that the DPD update time is significantly larger than the polynomial ones (i.e., 2.5 orders of magnitude higher, see C3-C4 vs C6-C11). This is clearly an obstacle for the adoption of ANN-based MIMO DPD schemes unless dimensionality reduction is combined with significant training data length reduction factors (and together with efficiently parallelized ANN-specific processing HW). The time taken to train the ANN at the first MIMO DPD iteration is in the order of magnitude (between 1 and 3 times higher) of the time needed to train the MIMO DPD in the remaining iterations with different data and PAPR statistics MIMO waveforms until the desired performance is reached (i.e., typically 7-8 iterations are needed). This is due to the fact that in the first iteration the DPD coefficients are trained from scratch, and it takes longer to find the coefficients that minimize the ANN cost function, and the following training iterations benefit from using the previously calculated and updated coefficients when training with new data. Having pre-trained ANNs can also contribute to reduce the overall MIMO DPD training time (TT in Table I).

In Table II, the dimensionality reduction techniques are preceded by the training data length reduction techniques as per Section V, and the length reduction processing time is both identified and further added to the total DPD update time count. In the MISO POLY scheme, the best data length reduction techniques are SBS and UES, but when combined with OLS, UES performs about 1.5 dB NMSE and 1.2 dB ACPR better and is slightly faster (C14 vs. C13). MeS does not provide acceptable ACPR performance when having strong cross couplings for any of the DPD architectures (the better performing case is shown for the MISO POLY architecture in C12). In the MISO POLY case, applying OLS reduction by 8 is already a powerful tool to significantly shorten the training time. When applying UES-OLS additional 35% DPD update time reduction can be obtained paying very little performance cost with regards to applying only OLS (C4 vs. C14).

With the ANN-based schemes, when having very strong cross couplings it is hard to reach the desirable performance when applying data length reduction right at the first DPD training iteration where the ANN coefficients are calculated from scratch. To avoid NMSE and ACPR losses of about 1-1.5 dB and 2-2.5 dB (C16-C17 vs. C19-C20 and C22-23 vs. C25-C26), respectively, pretrained first iteration coefficients that have been calculated without training data length reduction can be loaded, and the length reduction be applied starting at the second iteration. This is clearly what better works for such strong crosstalk conditions but, if pretrained ANN coefficients

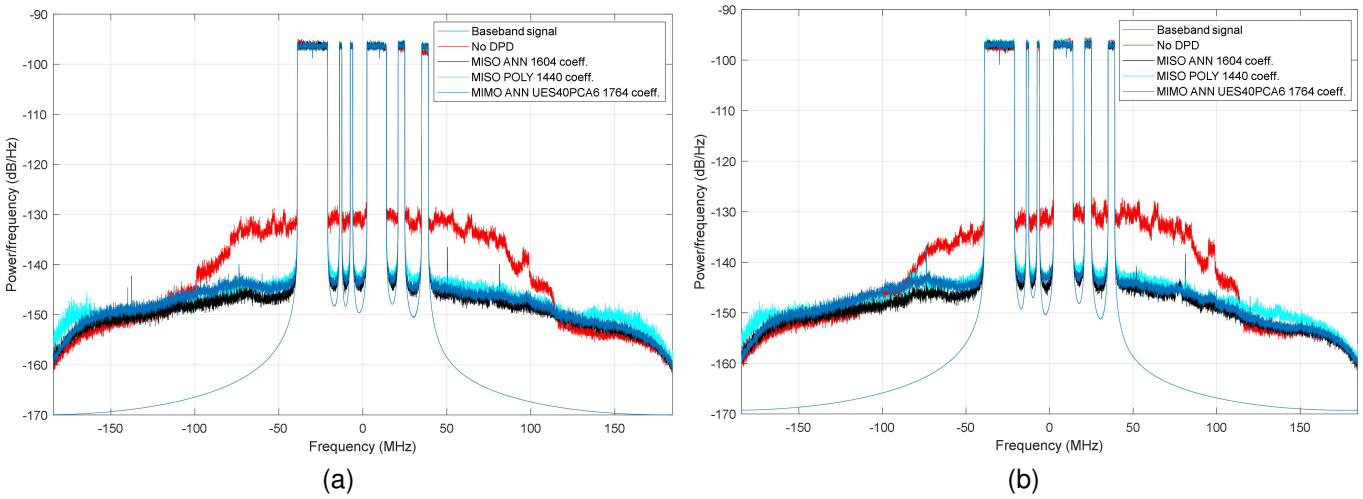

Fig. 12: Power spectral density plots for (a) CH1 and (b) CH2 before and after DPD linearization, considering the GMPNLC polynomial architecture (C2), the MISO ANN architecture with OLS reduction by 16 (C7), and the MIMO ANN architecture with UES reduction by 40 and PCA reduction by 6 (C23).

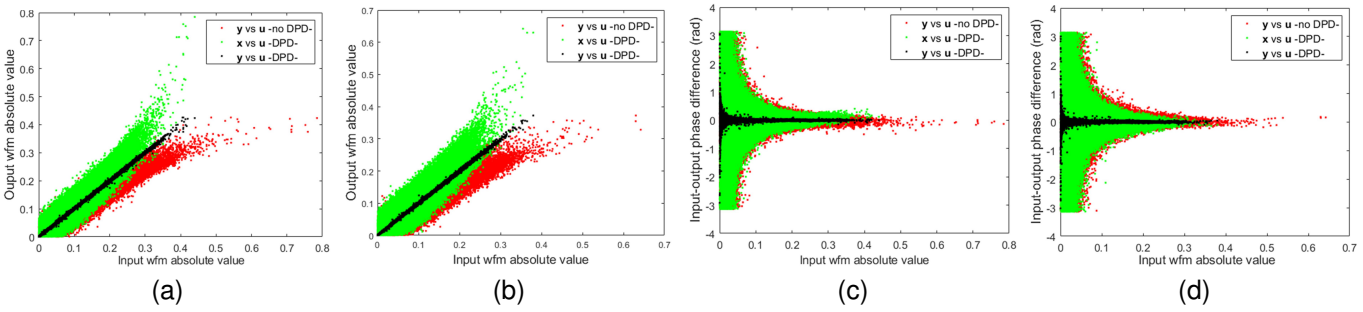

Fig. 13: AM-AM plots for (a) CH1 and (b) CH2 and AM-PM plots for (c) CH1 and (d) CH2 before and after DPD linearization, considering the MISO ANN architecture with OLS reduction by 16 (C7).

TABLE II: Training Data Length Reduction Combined with Dimensionality Reduction Benchmark for MIMO DPD.

	CASE. METHOD	RED. FACT.	DPD COEFF.	NMSE CH2 [dB]	NMSE CH1 [dB]	ACPR CH2 [dBc]	ACPR CH1 [dBc]	LENGTH REDUX TIME [s]	DPD TRAIN. TIME [s]	TOT. DPD UPD. TIME [s]
	C12. MeS-OLS	10_8	180	-32.6	-33.6	-39.7	-40.5	238.6	6.7	245.3
MISO POLY	C13. SBS-OLS	20_8	180	-35.2	-35.7	-43.1	-43.2	50	3.3	53.3
	C14. UES-OLS	**10_8**	**180**	**-36.8**	**-36.9**	**-44.4**	**-44.5**	**42.7**	**7.1**	**49.8**
MISO ANN 1^{st} iter pretrained	C15. CBS-PCA	80_4	3604	-36.2	-36.8	-43.5	-43.6	68.3	146.4	214.7
	C16. UES-PCA	**80_4**	**3604**	**-36.2**	**-37.2**	**-45**	**-44.8**	**12.9**	**92**	**104.9**
	C17. UES-PCA	**40_4**	**3604**	**-36.8**	**-37.7**	**-45.4**	**-45.7**	**13**	**170.6**	**183.6**
	C18. UES-OLS	80_16	1604	-33.5	-35.3	-43.8	-44.5	12.9	67.5	80.4
MISO ANN	C19. UES-PCA	80_4	3604	-35.5	-35.7	-43.1	-42.1	15.1	211.9	227
	C20. UES-PCA	40_4	3604	-36.2	-37.1	-44.6	-44.6	13.3	242.3	255.6
MIMO ANN 1^{st} iter pretrained	C21. CBS-PCA	60_6	1764	-36.4	-37.6	-45.4	-45.8	45	265	310
	C22. UES-PCA	**60_6**	**1764**	**-36.2**	**-37.3**	**-44.9**	**-44.7**	**15.5**	**93.5**	**109**
	C23. UES-PCA	**40_6**	**1764**	**-36.4**	**-37.6**	**-45.6**	**-45.9**	**15.2**	**200.3**	**215.5**
	C24. UES-OLS	40_16	964	-36.1	-36.6	-43.1	-43.8	15.3	87.8	103.1
MIMO ANN	C25. UES-PCA	60_6	1764	-35.2	-36	-43	-43.1	17.7	258	275.7
	C26. UES-PCA	40_6	1764	-35.9	-36.8	-45	-44.4	17.6	534.2	551.8

are not available, UES can be also applied from the first DPD iteration by decreasing the UES training data length reduction factor before calculating a new set of coefficients (C16 vs. C20 and C22 vs. C26). When compared to the optimal pretrained case, the performance will be similar and the training time will be kept in the same order of magnitude (i.e., 2-4 times higher). With ANNs, and when precombined with PCA, CBS works better than SBS. While in the MISO ANN UES shows 1-1.5 dB better ACPR performance and is about 1.5 times faster than CBS (C15 vs. C16), in the MIMO ANN scheme CBS may perform similarly or even better than UES, but the DPD updating time of the former can be 2-3 times larger (C21 vs. C22). Finally, while combining UES and OLS with a high dimensionality reduction factor can be about 2 times faster

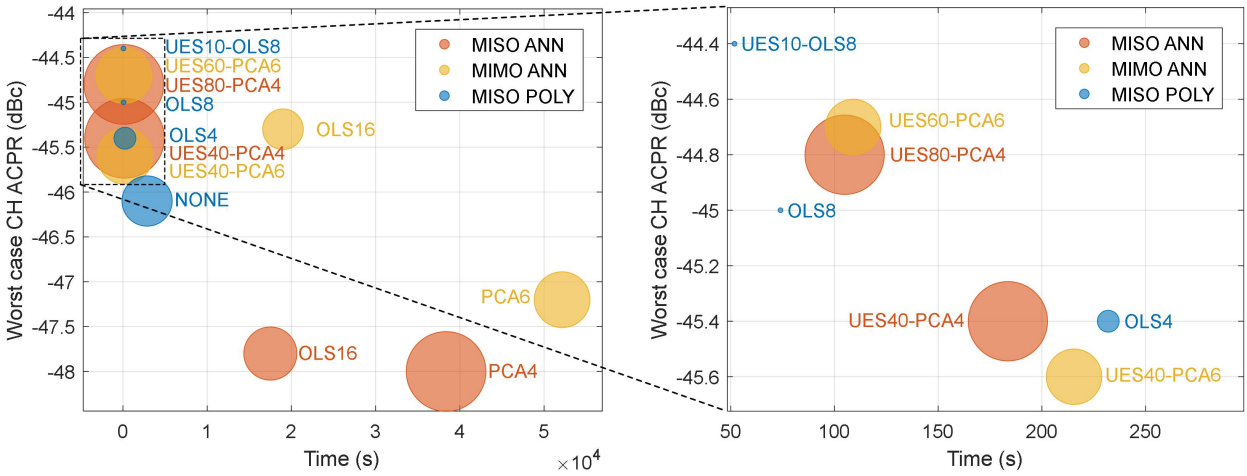

Fig. 14: Bubble plot with time versus worst case channel ACPR and number of coefficients (i.e., proportional to bubble size) considering the three presented MIMO DPD architectures.

than using UES-PCA with a lower dimensionality reduction factor (i.e., comparing 1^{st} iteration pretrained schemes), such advantage is not enough to compensate for the performance loss (C17 vs. C18 and C23 vs. C24).

At this point, it is important to remark that with the UES-PCA method being applied to the ANN-based MIMO DPD schemes, we can approach the polynomial MIMO DPD training times and we also have some degrees of freedom to overcome the linearization performance of the polynomial. One example may be found in Table II by comparing the MISO POLY case when applying OLS reduction by 4 with either the MISO ANN or the MIMO ANN cases applying UES-PCA (C4 vs. C17 and C23). The linearized spectra plots obtained when using the MIMO ANN scheme after applying UES-PCA in C23 are also added to Fig. 12 to show similar spectral contention when compared to the MISO POLY case without basis reduction in C2 (i.e., the comparison is done here for a similar number of coefficients, but featuring the MIMO-ANN scheme lower training time). The total DPD update time of the MISO/MIMO ANN schemes is reduced about 2 orders of magnitude when comparing the use of PCA or OLS with either UES-PCA and UES-OLS (C6-C11 vs. C15-C26). The gain with respect to the nominal ANNs without dimensionality reduction would be above three orders of magnitude. Applying dimensionality reduction to the herein ANN topologies (for 2x2 MIMO DPD) with factors between 6 and 16 deliver reductions in the number of ANN coefficients between the 55% and the 75%, respectively. Note that the higher the number of antennas, the more impact that the dimensionality reduction will have over the total number of ANN coefficients (the weights at the first hidden layer will increase far more significantly than the weights and biases at others).

The most representative results combining different MIMO DPD architectures with the proposed techniques are also summarized in Fig. 14. In general terms, it is evident that according to our 2x2 test scenario the ANN-based DPD schemes may feature higher complexity than the polynomial ones, but also similar adaptivity speed and better performance. Such benefits could even be more evident when considering larger arrays where the number of MIMO DPD coefficients may increase with the number of antennas much faster in the polynomial approach than in ANNs (i.e., MIMO ANN scheme).

VII. Conclusion

This work has provided an insight on training data selection and dimensionality reduction techniques for faster adaptivity and more efficient polynomial and ANN MIMO predistorters. The MIMO ANN DPD schemes can outperform the polynomial ones in complex scenarios but show unaffordable training times. OLS and PCA allow reducing the number of DPD coefficients and training time but need to be combined with data length reduction techniques for fast adaptivity. The batch selection techniques and the UES method contribute to significantly reduce the ANN DPD training time and overcome the limitations of histogram-based data selection mechanisms under strong cross couplings. This work contributes with two novel strategies: the first combines UES with OLS for polynomial MIMO predistorters, and the second combines UES with PCA for ANN schemes. With these techniques we have reduced the MIMO DPD training times by two orders of magnitude, the number of coefficients by a factor between 2 and 10 (depending on the MIMO DPD architecture), and meet the required performance for a worst-case scenario.

Acknowledgment

The authors would like to thank Dr José A. García and Dr Nieves Ruiz from the Universidad de Cantabria for the donation of the two wonderful class-J PA used in the experiments.

References

[1] 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.

[2] K. Hausmair, S. Gustafsson, C. Sanchez-Perez, P. N. Landin, U. Gustavsson, T. Eriksson, and C. Fager, "Prediction of nonlinear distortion in wideband active antenna arrays," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 11, pp. 4550–4563, Nov. 2017.

[3] P. Handel and D. Ronnow, "MIMO and massive MIMO transmitter crosstalk," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1882–1893, Mar. 2020.

[4] P. Handel, O. T. Demir, E. Bjornson, and D. Ronnow, "Impact of backward crosstalk in 2 × 2 MIMO transmitters on NMSE and spectral efficiency," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4277–4292, Jul. 2020.

[5] P. L. Gilabert, D. López-Bueno, T. Quynh Anh Pham, and G. Montoro, "Machine learning for digital front-end," in *Machine Learning for Future Wireless Communications*. John Wiley & Sons, Ltd, 2020, ch. 17, pp. 327–381.

[6] S. Bassam, M. Helaoui, and F. Ghannouchi, "Crossover digital predistorter for the compensation of crosstalk and nonlinearity in MIMO transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 57, no. 5, pp. 1119–1128, May 2009.

[7] 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.

[8] D. Saffar, N. Boulejfen, F. M. Ghannouchi, A. Gharsallah, and M. Helaoui, "Behavioral modeling of MIMO nonlinear systems with multivariable polynomials," *IEEE Trans. Microw. Theory Techn.*, vol. 59, no. 11, pp. 2994–3003, Nov. 2011.

[9] L. Anttila, P. Handel, and M. Valkama, "Joint mitigation of power amplifier and I/Q modulator impairments in broadband direct-conversion transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 4, pp. 730–739, Apr. 2010.

[10] S. Amin, P. N. Landin, P. Handel, and D. Ronnow, "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.

[11] Z. A. Khan, E. Zenteno, P. Handel, 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.

[12] 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.

[13] R. Zayani, R. Bouallegue, and D. Roviras, "Crossover neural network predistorter for the compensation of crosstalk and nonlinearity in MIMO OFDM systems," in *Proc. IEEE Int. Symp. On Personal Indoor and Mobile Radio Commun. (PIMRC)*, Sep. 2010, pp. 966–970.

[14] P. Jaraut, M. Rawat, and F. M. Ghannouchi, "Composite neural network digital predistortion model for joint mitigation of crosstalk, I/Q imbalance, nonlinearity in MIMO transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 11, pp. 5011–5020, Nov. 2018.

[15] D. Wang, M. Aziz, M. Helaoui, and F. M. Ghannouchi, "Augmented real-valued time-delay neural network for compensation of distortions and impairments in wireless transmitters," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 1, pp. 242–254, Jan. 2019.

[16] C. Yu, J. Jing, H. Shao, Z. H. Jiang, P. Yan, X.-W. Zhu, W. Hong, and A. Zhu, "Full-angle digital predistortion of 5G millimeter-wave massive MIMO transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 7, pp. 2847–2860, Jul. 2019.

[17] 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, May 2020.

[18] 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.

[19] Z. Wang, L. Guan, and R. Farrell, "Undersampling observation-based compact digital predistortion for single-chain multiband and wideband direct-to-RF transmitter," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 12, pp. 5274–5283, Dec. 2017.

[20] R. Zhu, "Gradient-based sampling: An adaptive importance sampling for least-squares," in *Proc. Int. Conf. on Neural Inf. Process. Syst.*, Dec. 2016, pp. 406–414.

[21] J. Kral, T. Gotthans, R. Marsalek, M. Harvanek, and M. Rupp, "On feedback sample selection methods allowing lightweight digital predistorter adaptation," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 67, no. 6, pp. 1976–1988, Jun. 2020.

[22] T. Wang and P. L. Gilabert, "Mesh-selecting for computational efficient PA behavioral modeling and DPD linearization," *IEEE Microw. Wireless Compon. Lett.*, vol. 31, no. 1, pp. 37–40, Jan. 2021.

[23] D. Lopez-Bueno, P. L. Gilabert, and G. Montoro, "Dataset reduction for neural network based digital predistorters under strong nonlinearities," in *Proc. IEEE Topical Conf. RF/Microw. Power Modeling Radio Wireless Appl. (PAWR)*, Jan. 2021, pp. 8–11.

[24] A. Barry, W. Li, J. A. Becerra, and P. L. Gilabert, "Comparison of feature selection techniques for power amplifier behavioral modeling and digital predistortion linearization," *Sensors*, vol. 21, no. 17, pp. 1–28, Aug. 2021.

[25] Q. A. Pham, G. Montoro, D. Lopez-Bueno, and P. L. Gilabert, "Dynamic selection and estimation of the digital predistorter parameters for power amplifier linearization," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 10, pp. 3996–4004, Oct. 2019.

[26] M. Bartlett, J. Movellan, and T. Sejnowski, "Face recognition by independent component analysis," *IEEE Trans. Neural Netw.*, vol. 13, no. 6, pp. 1450–1464, Nov. 2002.

[27] S. Chen, C. Cowan, and P. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Trans. Neural Netw.*, vol. 2, no. 2, pp. 302–309, Mar. 1991.

[28] G. Huang, S. Song, and C. Wu, "Orthogonal least squares algorithm for training cascade neural networks," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 59, no. 11, pp. 2629–2637, Nov. 2012.

[29] G. Yang, F. Liu, H. Li, W. Qiao, C. Jiang, and L. Su, "Sample selection method for digital predistortion with incremental dimension of coefficients," in *Proc. IEEE MTT-S Int. Wireless Symp. (IWS)*, May 2021, pp. 1–3.

[30] E. Zenteno, S. Amin, M. Isaksson, D. Ronnow, and P. Handel, "Combating the dimensionality of nonlinear MIMO amplifier predistortion by basis pursuit," in *Proc. European Microw. Conf. (EuMC)*, Oct. 2014, pp. 833–836.

[31] A. Brihuega, M. Abdelaziz, L. Anttila, M. Turunen, M. Allen, T. Eriksson, and M. Valkama, "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.

[32] X. Wang, Y. Li, H. Yin, C. Yu, Z. Yu, W. Hong, and A. Zhu, "Digital predistortion of 5G multiuser MIMO transmitters using low-dimensional feature-based model generation," *IEEE Trans. Microw. Theory Techn.*, vol. 70, no. 3, pp. 1509–1520, Mar. 2021.

[33] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *Int. J. Control*, vol. 50, no. 5, pp. 1873–1896, Nov. 1989.

[34] J. A. Becerra, M. J. Madero-Ayora, J. Reina-Tosina, C. Crespo-Cadenas, J. Garcia-Frias, and G. Arce, "A doubly orthogonal matching pursuit algorithm for sparse predistortion of power amplifiers," *IEEE Microw. Wireless Compon. Lett.*, vol. 28, no. 8, pp. 726–728, Aug. 2018.

[35] M. Hagan and M. Menhaj, "Training feedforward networks with the Marquardt algorithm," *IEEE Trans. Neural Netw.*, vol. 5, no. 6, pp. 989–993, Nov. 1994.

[36] J. Reina-Tosina, M. Allegue-Martínez, C. Crespo-Cadenas, C. Yu, and S. Cruces, "Behavioral modeling and predistortion of power amplifiers under sparsity hypothesis," *IEEE Trans. on Microw. Theory and Tech.*, vol. 63, no. 2, pp. 745–753, Feb. 2015.

[37] P. L. Gilabert, G. Montoro, D. López, N. Bartzoudis, E. Bertran, M. Payaro, and A. Hourtane, "Order reduction of wideband digital predistorters using principal component analysis," in *Proc. IEEE MTT-S Int. Microw. Symp. Dig. (IMS)*, Jun. 2013, pp. 1–4.

[38] D. López-Bueno, Q. A. Pham, G. Montoro, and P. L. Gilabert, "Independent digital predistortion parameters estimation using adaptive principal component analysis," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 12, pp. 5771–5779, Dec. 2018.