On Feedback Sample Selection Methods Allowing Lightweight Digital Predistorter Adaptation

Jan Kral, Tomas Gotthans, *Member, IEEE*, Roman Marsalek, *Member, IEEE*, Michal Harvanek, and Markus Rupp, *Fellow, IEEE*

***Abstract*— In modern communication systems advanced techniques such as digital predistortion (DPD) are required to satisfy stringent demands on transmitter linearity and efficiency. DPD, however, increases the hardware and computational complexity of transmitters. In this article we show that the computational complexity of DPD adaptation can be drastically reduced if a low number of samples is properly selected for DPD adaptation. For this purpose we propose methods of sample selection for DPD adaptation. Among the proposed methods, the highest computational complexity reduction is achieved by a method based on the histogram of signal magnitudes, optimised with respect to characteristics of the radio frequency power amplifier and of the transmitted signal. Simulations indicate that this proposed method can reduce the computational complexity of DPD adaptation by a factor of up to 400 while the linearisation performance of conventional methods is preserved. Besides simulations for three models of distinct power amplifiers, measurements on a real power amplifier further verify the linearisation capabilities of the proposed methods.**

***Index Terms*— Power amplifier linearisation, digital predistortion, low complexity adaptation, histogram-based sample selection, sample subset reduction, direct learning architecture.**

I. Introduction

MODERN communication systems provide ever-increasing data throughput, putting stringent demands on transmitters and receivers, especially on hardware imperfections. In the currently developed fifth generation (5G) systems, one of the main issues is the linearity of radio frequency (RF) power amplifiers (PA) in transmitters [1]. A widely used technique to improve transmitter linearity is digital predistortion (DPD). DPD introduces an artificial nonlinearity with complementary characteristics to cancel the PA nonlinearity.

Manuscript received September 29, 2019; revised January 16, 2020; accepted February 6, 2020. Date of publication February 28, 2020; date of current version June 3, 2020. This work was supported by the Czech Science Foundation (GACR) under Project 17-18675S Future transceiver techniques for the society in motion and in part by the Czech Ministry of Education in a Frame of the National Sustainability Program under Grant LO1401 INWITE. This article was recommended by Associate Editor H. Sjoland. *(Corresponding author: Jan Kral.)*

Jan Kral, Tomas Gotthans, Roman Marsalek, and Michal Harvanek are with the Department of Radio Electronics, Brno University of Technology, 616 00 Brno, Czech Republic (e-mail: jan.kral@vutbr.cz; gotthans@feec.vutbr.cz; marsaler@feec.vutbr.cz).

Markus Rupp is with the Department of Radio Electronics, Brno University of Technology, 616 00 Brno, Czech Republic, and also with the Institute of Telecommunications, TU Wien, 1040 Vienna, Austria.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSI.2020.2975532

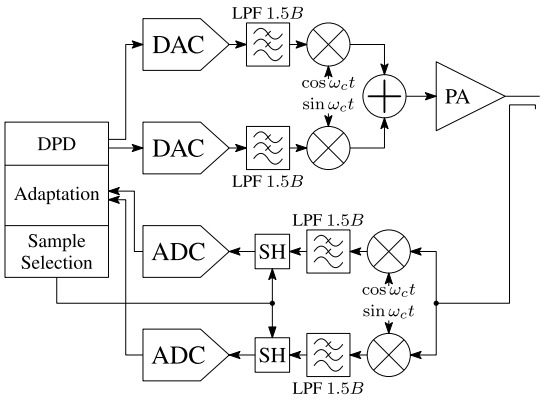

Fig. 1. Linearisation of an RF PA with the DPD using an optional sample selection method. In the direct path, there are digital to analogue converters (DAC) accompanied by reconstruction lowpass filters (LPF) with bandwidth $1.5B$, where B is the desired channel bandwidth. In the feedback, there are LPFs with bandwidth $> 1.5B$ to limit the noise bandwidth, optional sample and hold (SH) circuits and ADCs. Note that the resulting bandwidth of the forward and feedback paths is $3B$ due to IQ sampling. Although the depicted S/H circuits as discrete components are one possible approach, a more practical implementation would employ ADCs with integrated S/H circuits and with periodic and equidistant sampling, performing a sample selection from a sample buffer in the digital domain.

The required characterisation of PA nonlinearity is achieved by using a feedback from the PA output. Knowing the desired output and the actual output, one can induce a correction term which cancels out the PA error. The block diagram of a transmitter linearised by DPD is depicted in Fig. 1. In comparison to a transmitter without DPD, there are several additional components in the feedback path. The two analogue-to-digital converters (ADC) sampling in-phase and quadrature (IQ) signals are the most power hungry of them. Researchers often attempt to reduce the complexity of the DPD-enabled hardware realisation. One possible approach is the sampling of the in-phase signal, also called real-valued feedback, as presented in [2], [3].

Another approach to relaxing the feedback requirements is to sample the feedback signals with low sample rate ADCs making the feedback signals undersampled. This topic has been studied by Liu *et al.* in [4], by Zhang and Feng in [5], and by Wang *et al.* in [6]. In [7] the authors first trained a PA forward model and then applied it for DPD adaptation by a standard indirect learning architecture (ILA) [8]. An improved version of DPD with undersampled feedback was presented in [9]–[11], where the authors utilised

a direct learning architecture (DLA) [8] to avoid the calculation of the PA forward model. Zhang *et al.* in [12] and Guan *et al.* in [13] combined real-valued feedback and undersampled feedback methods and presented DPD adaptation with a single undersampling ADC. Hammler *et al.* in [14] chose a different approach and proposed DPD with a special feedback sampling in the frequency domain which allowed them to reduce the acquisition bandwidth, sample rate, and required number of processed samples.

Along with the effort to relax the demands on hardware, there is also an objective to decrease computational complexity of DPD adaptation [15]–[17]. Benedict and Honcharenko in [18], [19] showed a general idea for DPD adaptation by a subset of selected samples based on the probability distribution functions of the signals used. In a different manner, the probability distribution functions were used for DPD adaptation in [20], where the authors transformed the distribution function of the input into the distribution function of the output to identify the PA model. In [21] the authors introduced a method for model-order reduction based on principal component analysis theory, and a mesh-selecting method for reducing the number of required equations. Wang *et al.* in [22] proposed a selection of compressed training data for the extracting an RF PA behavioural model. They selected samples of signal magnitudes constituting a scaled histogram with similar statistical properties. In their work, however, they did not consider the PA characteristics.

In this paper, we propose methods for the selection of samples for DPD adaptation allowing computational complexity reduction: a method based on the identification of important samples using QR decomposition [23], a gradient-based sampling method [24], and two histogram-based methods. The first histogram method equalises the histogram of signal magnitudes to ensure evenly sampled PA characteristics while the second one optimises a histogram optimised by a genetic algorithm, which respects both the transmitted signal statistical properties and the specific PA characteristics. All these methods are analysed with respect to a conventional DPD and a simple undersampling method [10], [11], [13]. Moreover, we provide a detailed analysis of the computational complexity of these methods and show how significantly some of them can reduce the required computational resources. The performance of the proposed sample selection methods is shown in simulations and eventually verified by measurements.

The main contributions of this paper are

- proposal of methods for sample selection for DPD adaptation compatible with the conventional DPD-enabled transmitter architectures, especially the histogram-based method with the histogram optimised by a genetic algorithm, which highly reduces the computational complexity while preserving the linearisation capabilities,
- comparison of the computational complexity of proposed methods,
- availability of all MATLAB source codes[1] for reproducing the simulation results.

[1] www.github.com/jankralx/dpd_sample_selection

The rest of the paper is organised as follows. We introduce the theory behind the sample selection for DPD adaptation in Sec. II. Sample selection methods are proposed and discussed in Sec. III. We show how the proposed sample selection methods reduce the computational and hardware resources needed for DPD adaptation in Sec. IV and Sec. IV-D. Finally, simulations are described in Sec. V and measurements in Sec. VI.

II. DPD Adaptation With Undersampled Feedback

A. PA Model Identification With Undersampled Feedback

For easier understanding we first present the key principle of our methods on the example of the identification of the PA model, and later we extend the same principle to the identification of the DPD model. The identification of the PA model is closely related to DPD and is a first step in some linearisation methods [25], [26].

Let us assume that DPD is modelled by a baseband memory polynomial (MP) model [27]. We have chosen the MP model for its formal simplicity, but the following concepts can be fully evolved for any model which is linear in its unknown parameters, e.g. all models derived from full Volterra series such as Generalized Memory Polynomial (GMP) model [28], the simplified 2nd-order dynamic deviation reduction model (DDR2) Volterra series [29], etc.

The discrete baseband output sample $y(n)$ of the PA model is given as [27]

$$y(n) = \sum_{k=1}^{K} \sum_{q=0}^{Q} b_{k,q} x(n-q)|x(n-q)|^{k-1}, \tag{1}$$

where n is the sample index, x is the PA input, $b_{k,q}$ is a coefficient of the PA model, and K and Q represent the maximum model nonlinearity order and memory length respectively. The product $x(n-q)|x(n-q)|^{k-1}$ is often called the basis waveform, the basis function, or a kernel. We denote it as

$$\phi_{k,q}^{(x)}(n) = x(n-q)|x(n-q)|^{k-1}. \tag{2}$$

To identify the coefficients $b_{k,q}$ of the PA model, conventional methods construct a system of equations taking consecutive input and output samples of the PA and insert them into (1). Here we would like to point out that it is not necessary to take consecutive samples of the PA output $y(n)$ to solve the model coefficients $b_{k,q}$, we can rather take the PA output on the left hand side of (1) with arbitrary sample indices $n_i \in \mathbb{N}; i \in \{1, 2, \ldots, N\}$ that satisfy the condition $n_i \neq n_\kappa; \forall i, \kappa : i \neq \kappa$. The constructed equation system can be expressed in the matrix form by arranging the output samples, model coefficients and basis functions into vectors, i.e.

$$\begin{aligned}
\phi_{k,q}^{(x)} &= \begin{bmatrix} \phi_{k,q}^{(x)}(n_1) & \phi_{k,q}^{(x)}(n_2) & \ldots & \phi_{k,q}^{(x)}(n_N) \end{bmatrix}^T, \\
\mathbf{y} &= \begin{bmatrix} y(n_1) & y(n_2) & \ldots & y(n_N) \end{bmatrix}^T, \\
\mathbf{b} &= \begin{bmatrix} b_{1,0} & b_{1,1} & \ldots & b_{1,Q} & b_{2,0} & \ldots & b_{K,Q} \end{bmatrix}^T, \\
\mathbf{U}_x &= \begin{bmatrix} \phi_{1,0}^{(x)} & \phi_{1,1}^{(x)} & \ldots & \phi_{1,Q}^{(x)} & \phi_{2,0}^{(x)} & \ldots & \phi_{K,Q}^{(x)} \end{bmatrix},
\end{aligned}$$

where $\mathbf{b}$ is a column vector with $P = K(Q+1)$ rows, and the size of the matrix $\mathbf{U}_x$ is $N \times P$. The equation system is then expressed as

$$\mathbf{y} = \mathbf{U}_x \mathbf{b}. \tag{3}$$

By solving (3) we obtain the model coefficients $\mathbf{b}$ and hence identify the PA model. Typically we choose $N \gg P$ to sufficiently sample the whole characteristics of the PA with diverse signal samples, and to mitigate the influence of noise and other imperfections of a real system. The solution is usually given as a projection of $\mathbf{b}$ into the column space of $\mathbf{U}_x$, in other words it is the least squares (LS) solution which minimises the difference between the observed PA output and the model output

$$\mathbf{b} = (\mathbf{U}_x^H \mathbf{U}_x)^{-1} \mathbf{U}_x^H \mathbf{y}. \tag{4}$$

We can imagine the above procedure such that we first take consecutive samples of the PA output, build a conventional overdetermined equation system and before solving it, we leave out some arbitrary rows from the matrix $\mathbf{U}_x$ and the equivalent samples from the vector $\mathbf{y}$. We would like to emphasise that the reduction of the equation system preserves the memory modelling, because (1) is still fully valid. Please notice that only the PA output $y(n)$, left hand side of the system in (3), can be arbitrarily undersampled. The PA input $x(n)$ has to be known, in order to cover at least the memory modelling of the PA to build the matrix $\mathbf{U}_x$.

Hereinafter the described method is referred to as the sample selection method (SSM). In the following sections we show some possible ways of employing SSM for DPD identification.

B. Direct Learning Architecture

DLA [8] is an iterative method which directly solves $G(B(z)) = y$ with G being a transfer function of PA and B a transfer function of DPD. The solution $B(z) = G^{-1}(y)$ is a nonlinear problem and can be solved by the damped Newton's method, which can be defined for DPD as [30]

$$\mathbf{b}'(m+1) = \mathbf{b}'(m) - \mu \mathbf{e}(m), \tag{5}$$

where m is the iteration index, $\mathbf{b}'(m)$ and $\mathbf{b}'(m+1)$ are DPD coefficients in the m-th and $(m+1)$-th iteration, μ is the iteration step size, and $\mathbf{e}(m)$ is the coefficients error vector for the m-th iteration. It has the same dimensions as the coefficient vector $\mathbf{b}'$ and is given as the LS solution of

$$\Delta = \mathbf{U}_z \mathbf{e}, \tag{6}$$

where $\Delta = \mathbf{z} - \mathbf{y}$ and $\mathbf{z}$ is the desired scaled PA output and the DPD input.

We assume that (6) is originally overdetermined, therefore we can omit arbitrary rows of the matrix $\mathbf{U}_z$ and the equivalent samples of the vector Δ, in the same way as in Sec. II-A. They can be denoted as

$$\mathbf{z} = \begin{bmatrix} z(n_1) & z(n_2) & \dots & z(n_N) \end{bmatrix}^T,$$
$$\mathbf{U}_z = \begin{bmatrix} \phi_{1,0}^{(z)} & \phi_{1,1}^{(z)} & \dots & \phi_{1,Q}^{(z)} & \phi_{2,0}^{(z)} & \dots & \phi_{K,Q}^{(z)} \end{bmatrix}.$$

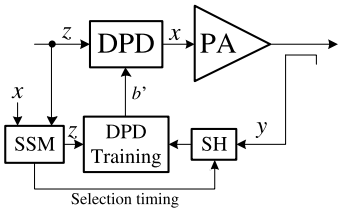

Fig. 2. Block diagram of the transmitter with a baseband PA model linearised by DPD with DLA. The PA output is sampled by the sample-and-hold circuit (SH) at times driven by SSM based on the PA input samples x.

The notation of the final solution of (5) using SSM does not differ from the conventional DLA [8]

$$\mathbf{b}'(m+1) = \mathbf{b}'(m) - \mu (\mathbf{U}_z^H \mathbf{U}_z)^{-1} \mathbf{U}_z^H (\mathbf{z} - \mathbf{y}). \tag{7}$$

The block diagram of DLA is depicted in Fig. 2. Please notice that SSM allows arbitrary undersampling of only the PA output y. The PA output is required to be sampled with a sufficient bandwidth, which is required for all the undersampling methods [10], [11], [13]. The desired scaled PA output z has to be sufficiently known to construct the matrix $\mathbf{U}_z$. This usually does not represent any complication as the desired scaled PA output is also the DPD input.

Although there are options how to employ SSMs with ILA, in this paper we focus our analysis and findings only on DLA, because we find DLA more efficient and usable than ILA-based methods, despite its disadvantages such as the initial solution, stability issues, etc.

III. Sample Selection Methods

The best DPD adaptation can be achieved if the observation errors of the input and output are uncorrelated [15]. However, the adjacent samples used by the conventional DPD are not independent and therefore the observation errors are correlated. To minimise the observation errors, the LS estimator requires a high number of samples N, usually $N > 1000$ [15]. For a small N, using subsequent samples causes the system of equations to be ill-conditioned. Furthermore, a limited number of subsequent samples cannot cover the statistical properties of the transmitted signal.

We show that the proposed SSM does not suffer from the mentioned drawbacks for small N if the samples $n_1, n_2, \dots, n_N$ are selected carefully. The problem for the sample selection method can be defined as the selection of N samples from all samples which were acquired by the feedback ADCs. The number of all acquired samples is N_0 and corresponds to the acquisition time and hence to the required update rate of DPD coefficients. The number of selected samples is naturally limited by the condition $N < N_0$.

A. Undersampling

One possible approach to SSM is simple undersampling of the feedback. It can bee seen as untargeted sample selection and henceforth it is referred to as undersampling-based sample selection (US). Similar approaches were presented in [10], [11], [13]. Although this method is very simple

and does not consider the statistics of the transmitted signal, it can improve the conditioning of the equation system, as it takes distant samples which have the potential to be more independent. Therefore this simple approach can be sufficient for less demanding applications.

B. Sample Selection Based on QR Decomposition

The problem of selecting the samples $n_1, n_2, \ldots, n_N$ for the calculation of new DPD coefficients can be solved analytically using QR decomposition with column pivoting [23], henceforth referred to as QR-based sample selection (QRS). First we construct the matrix $\mathbf{U}_z$ using all the samples from the feedback. Afterwards we perform the QR decomposition with column pivoting of the transposed matrix $\mathbf{U}_z$, which is defined as

$$\mathbf{U}_z^T \mathbf{E} = \mathbf{QR}, \tag{8}$$

where $\mathbf{R}$ is an $N_0 \times P$ upper triangular matrix; $\mathbf{Q}$ is an $N_0 \times N_0$ unitary matrix; and $\mathbf{E}$ is an $N_0 \times N_0$ permutation matrix. The first N columns of the permutation matrix identify the N most important rows of matrix $\mathbf{U}_z$ for the solution of new DPD coefficients. Formally, we can write that the sample indexes of selected samples are given as

$$\left[n_1, n_2, \ldots, n_N\right] = \left[1, 2, \ldots, N_0\right] \mathbf{E}_{1:N}, \tag{9}$$

where $\mathbf{E}_{1:N}$ is a submatrix of matrix $\mathbf{E}$ consisting of its first N columns.

A drawback of this method is that it improves the conditioning of the pseudoinverse of matrix $\mathbf{U}_z$, but this does not guarantee improving the DPD linearisation performance, because this method does neither respect vector Δ in the calculation of the DPD coefficients in (7) nor the PA characteristics.

Moreover, the practical usability of this method for real-time DPD adaptation is limited, because it requires performing the QR decomposition of the full matrix $\mathbf{U}$, which in principle already solves the solution using all feedback samples. Potential usage is limited to cases where a fast QR decomposition with low precision [31] is applied to identify a small number of important samples and thus requires fewer computation resources, and high precision calculation is performed by only the selected samples.

C. Gradient Sampling

Another method for sample selection can be the gradient-based sampling (GS) [24], henceforth referred to as GS-based sample selection (GSS). GS is a representative of the methods for solving large sample size least squares problems. For DLA, we calculate the gradient values g_i for all rows of matrix $\mathbf{U}_z$ and all samples of vector Δ [24]

$$g_i = ||\mathbf{u}_i(\Delta_i - \mathbf{u}_i \mathbf{e}_0)||, \tag{10}$$

where $||.||$ is the l_2-norm, $\mathbf{u}_i$ is the i-th row of matrix $\mathbf{U}_z$, Δ_i is the i-th element of vector Δ, and $\mathbf{e}_0$ is a given pilot estimate (good guess) for solving the coefficient error vector $\mathbf{e}$. Since $\mathbf{e}$ is ideally a zero vector in the converged state, we set $\mathbf{e}_0$ as a zero vector and thus we can simplify the calculation of gradient values to

$$g_i = ||\mathbf{u}_i \Delta_i||. \tag{11}$$

The probability that the i-th feedback sample and the equivalent row of matrix $\mathbf{U}_z$ are taken into calculation of the DPD adaptation in (7) is calculated as [24]

$$p_i = \frac{N g_i}{\sum_{i=1}^{N_0} g_i}. \tag{12}$$

Based on the calculated probabilities, N samples $n_1, n_2, \ldots, n_N$ are selected for the calculation of new DPD coefficients.

D. Histogram-Based SSMs

Even though both QRS and GSS are analytically based, they do not consider a priori knowledge of the PA characteristics nor signal statistics. To surpass this disadvantage, we propose an SSM based on the histogram of signal magnitudes of the PA input x. Let us make the following notations: J is the number of histogram bins, the target bin counts are denoted d_j with $j \in 1, 2, \ldots, J$, the lower and upper bin boundaries are θ_{j-1} and θ_j respectively (Fig. 3), D_j is a set of selected indices n_i whose samples belong to the j-th bin, defined as

$$D_j = \{n_i : \theta_{j-1} < |x(n_i)| < \theta_j\}, \tag{13}$$

and $\overline{\overline{D_j}}$ is the cardinality, the number of elements, of set D_j.

Samples for DPD adaptation are selected randomly such that the j-th histogram bin count reaches the target bin count d_j, i.e. cardinality $\overline{\overline{D_j}}$ is equal to d_j. At the same time, samples should be selected such that they are distant in time and hence the matrix conditioning is improved. Formally, the selected indices need to satisfy the condition

$$|n_i - n_\kappa| > \gamma\,;\ \forall i, \kappa : i \neq \kappa, \tag{14}$$

where γ is the minimum sample time distance and is set such that it ensures the independence of selected samples.

Even though the papers [21], [22] have proposed histogram methods with a slightly different usage, the condition of a minimum sample distance expressed by (14) has not, to the authors' best knowledge, been proposed for the histogram-based sample selection methods.

Evenly Distributed Histogram (EDH) is a possible setting for histogram-based sampling. If N is divisible by J, all target bin counts are set to the same constant $d_j = \frac{N}{J}$, otherwise they are selected randomly from the values $d_j \in \{\lfloor \frac{N}{J} \rfloor;\ \lfloor \frac{N}{J} \rfloor + 1\}$ such as $\sum d_j = N$. The motivation for making the histogram evenly distributed is to cover the whole amplitude/amplitude (AM/AM) characteristics of PA, as shown in Fig. 3. Unfortunately, this simple approach of setting target bin counts does not respect either the shape of the AM/AM characteristics of the used PA nor statistical properties of the transmitted signal (Fig. 3) which results into a DPD model with the similar modelling capabilities in all regions of AM/AM characteristics, even those not frequently utilised due to high peak-to-average

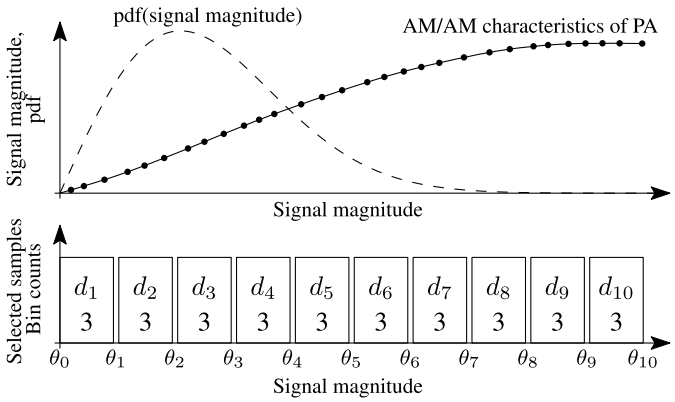

Fig. 3. Principle of the sample selection by EDH with $J = 10$ related to the probability density function (pdf) of signal magnitude and AM/AM characteristics of the PA. The depicted $N = 30$ selected samples are evenly distributed over the whole AM/AM characteristics of the PA, irrespective of the signal statistics nor the shape of AM/AM characteristics.

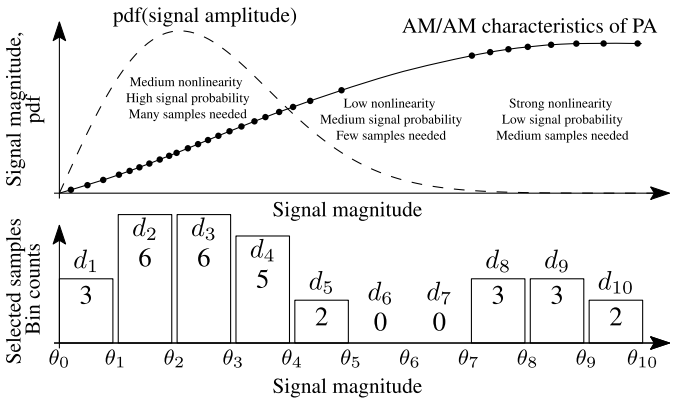

Fig. 4. Principle of the sample selection by GOH with $J = 10$ related to the probability density function (pdf) of signal magnitude and AM/AM characteristics of the PA. The depicted $N = 30$ selected samples are placed in regions with strong nonlinearity and/or high signal probability.

power ratio (PAPR) of the transmitted signal. In other words, if we consider a limited number of points for DPD adaptation, selecting points from regions that are highly nonlinear and with high signal probability provide more information for the DPD adaptation than selecting points from linear regions and low signal probability.

Genetically Optimised Histogram (GOH) is proposed to suppress EDH imperfections. An optimised histogram can respect the AM/AM characteristics of the PA and the statistics of the transmitted signal, as depicted in Fig. 4. We show that it is possible to set the target bin counts such that a chosen criterion is optimised. In general, the histogram bin counts should be set such that GOH selects samples from regions with a strong nonlinearity and a high signal amplitude probability.

In this paper, we optimise target histogram bin counts by the genetic algorithm [32] to minimise the normalised mean square error (NMSE) of the PA output with respect to the desired output. We assume that the histogram is optimised once for a specific type of PA and the transmit signals, just to set the target bin counts which are afterwards applied for DPD adaptation. Therefore, the higher computational cost of optimisation calculation can be neglected, although it may take several hours to optimise the histogram. Therefore, the optimisation is most likely unfeasible in real time in a transmitter, and we consider precalculated histograms only. However, we have not aimed to make the optimisation faster, which might be a topic for future research.

The application of GOH can be limited in cases of changing working conditions. The PA characteristics vary due to temperature changes and ageing. However, based on our experience, temperature variations cause mainly changes of the PA gain and only small differences of the shape of the AM/AM characteristics (usually within 10 %). A small margin in the number of selected samples N ensures a GOH insensitivity to these small changes. However, an extensive analysis have to be always done for a specific application.

Another potential limitation arises when the communication system employs signals with significantly changing statistics of the signal magnitudes. If a single precalculated histogram does not lead to sufficient linearisation performance, then we suggest to optimise the histogram for all different types of the signals and to apply the corresponding precalculated histogram in hardware.

In both cases, it is always possible to apply EDH which does not require the histogram optimisation. The simulations with results in Sec. V-G show that EDH requires an only slightly increased number of selected samples N or provides a little worse linearisation performance in comparison to GOH. GOH and EDH are further discussed in Sec. V-E and Sec. V-F.

IV. Reduction of Computational Complexity of DPD Adaptation

A small number of properly selected samples N significantly reduces the computational complexity of DPD adaptation. We evaluate the computational complexity of DPD adaptation with P coefficients and N samples with respect to the required number of real-valued multiplications $O_{\otimes}(N, P)$ and real-valued additions $O_{\oplus}(N, P)$. The following analysis does not include any optimisation and does not exploit the properties of the PA model, and the presented numbers of required multiplications and additions can therefore be seen as the upper bounds. We assume that a single complex multiplication requires four real-valued multiplications and three real-valued additions; a single complex addition requires two real-valued additions.

We split the analysis into three parts: calculation of the DPD model represented by the matrix $\mathbf{U}_z$, calculation of the coefficient error vector $\mathbf{e}$, and the coefficients update. The calculation of $\mathbf{U}_z$ for the MP model in (1) requires $O_{U\otimes}(N, P)$ real-valued multiplications and none of the real-valued additions. Assuming that the calculation of $\phi_{k,q}^{(x)}(n)$ is reused for the calculation of $\phi_{k+1,q}^{(x)}(n)$, the number of multiplications can be expressed as

$$O_{U\otimes}(N, P) \approx 2NP. \tag{15}$$

The calculation of coefficient error vector $\mathbf{e}$ in (7) requires $O_{\mathbf{e}\hat{\otimes}}(N, P)$ complex multiplications: $\frac{1}{2}NP(P + 1)$ for the

$\mathbf{U}_z^H\mathbf{U}_z$ matrix multiplication because the resulting matrix is of size $P\times P$ and is symmetrical, $O_{I\hat{\otimes}}(P)\approx P^3$ for the inversion of the symmetrical $P\times P$ complex matrix $(\mathbf{U}_z^H\mathbf{U}_z)$, NP for $(\mathbf{U}_z^H(\mathbf{z}-\mathbf{y}))$, and P^2 for the multiplication of the matrix inverse $(\mathbf{U}_z^H\mathbf{U}_z)^{-1}$ and the vector $\mathbf{U}_z^H(\mathbf{z}-\mathbf{y})$. The calculation of coefficient correction $\mathbf{e}$ in (7) requires $O_{\mathbf{e}\hat{\oplus}}(N,P)$ complex additions: $\frac{1}{2}(N-1)P(P+1)$ additions for $(\mathbf{U}_z^H\mathbf{U}_z)$ calculation, $O_{I\hat{\oplus}}(P)\approx P^3$ additions for the inversion of the symmetrical $P\times P$ complex matrix, N additions for subtraction $(\mathbf{z}-\mathbf{y})$, $(N-1)P$ for $\mathbf{U}_z^H(\mathbf{z}-\mathbf{y})$, and $(P-1)P$ additions for the multiplication of the matrix inverse $(\mathbf{U}_z^H\mathbf{U}_z)^{-1}$ and the vector $(\mathbf{U}_z^H(\mathbf{z}-\mathbf{y}))$. Complexity can be expressed as

$$O_{\mathbf{e}\hat{\otimes}}(N,P)\approx\frac{1}{2}NP(P+3)+P^2+P^3, \tag{16}$$

$$O_{\mathbf{e}\hat{\oplus}}(N,P)\approx\frac{1}{2}(N-1)P(P+3)+P^3+N+(P-1)P. \tag{17}$$

The calculation of a single DLA iteration in (5) additionally requires $2P$ real-valued multiplications and additions. The total number of real-valued multiplications and additions can therefore be estimated as

$$O_{\otimes}(N,P)=O_{U\otimes}(N,P)+4O_{\mathbf{e}\hat{\otimes}}(N,P)+2P\approx 2NP(P+4)+4P^2+4P^3+2P, \tag{18}$$

$$O_{\oplus}(N,P)=2O_{\mathbf{e}\hat{\oplus}}(N,P)+3O_{\mathbf{e}\hat{\otimes}}(N,P)+2P\approx\left(\frac{5}{2}NP-P\right)(P+3)+5P^2+5P^3+2N. \tag{19}$$

A. Additional Complexity of Histogram-Based SSMs

In terms of resources histogram-based SSMs are very simple, because they only require storing and incrementing the actual bin counts to select the samples. Hence additional computational resources required by these methods are negligible.

B. Additional Complexity of GSS

GSS additionally requires calculating the matrix $\mathbf{U}_z$ for all N_o samples which requires $O_{U\otimes}(N_0-N,P)\approx 2(N_0-N)P$ real-valued multiplications for the MP model. In (11) the calculation of $\mathbf{u}_i\Delta_i$ needs P complex multiplications while the calculation of the l_2-norm of a complex vector needs $5P$ real-valued multiplications and $4P-1$ real-valued additions, and this equation needs to be calculated N_0 times. Additional computational complexity to implement GSS is therefore

$$O_{\otimes}^{GS}(N,P,N_0)\approx 11N_0P-2NP, \tag{20}$$

$$O_{\oplus}^{GS}(N,P,N_0)\approx N_0(7P+1). \tag{21}$$

C. Additional Complexity of QRS

QRS also requires calculating the matrix $\mathbf{U}_z$ for all N_o samples and QR decomposition requires $33N_0+1$ complex multiplications and $8N_0$ complex additions [31]. Additional computational complexity to implement QRS is therefore

$$O_{\otimes}^{QR}(N,P,N_0)\approx 132N_0+2(N_0-N)P+4, \tag{22}$$

$$O_{\oplus}^{QR}(N,P,N_0)\approx 107N_0+3. \tag{23}$$

TABLE I
COMPARISON OF COMPUTATIONAL COMPLEXITY

		Conv. DPD	QRS	GSS	Hist. SSM
$P=14$	$O_{\otimes}$	DNC	$3.2\cdot10^6$	$3.1\cdot10^6$	$22\cdot10^3$
$N=20$	$O_{\oplus}$	DNC	$2.2\cdot10^6$	$2.0\cdot10^6$	$26\cdot10^3$
$P=14$	$O_{\otimes}$	DNC	$3.3\cdot10^6$	$3.1\cdot10^6$	$62\cdot10^3$
$N=100$	$O_{\oplus}$	DNC	$2.2\cdot10^6$	$2.1\cdot10^6$	$74\cdot10^3$
$P=14$	$O_{\otimes}$	$5.1\cdot10^6$	$8.0\cdot10^6$	$7.8\cdot10^6$	$5.1\cdot10^6$
$N=10^4$	$O_{\oplus}$	$6.0\cdot10^6$	$8.1\cdot10^6$	$7.9\cdot10^6$	$6.0\cdot10^6$

Conventional DPD does not converge (DNC) for a low number of selected samples N. The complexity in the table is calculated for $N_0=20\,000$ and the MP model with $K=7$, and $Q=1$.

D. Comparison of SSMs Complexity

In Table I we provide a comparison of computational complexity for the above SSMs in terms of the number of real-valued multiplications and additions. In all cases there are $N_0=20\,000$ samples collected from the feedback. The complexity is evaluated for the MP model with $K=7$, and $Q=1$ which leads to $P=14$ DPD coefficients. We can conclude that QRS is computationally the most demanding as it basically performs the inversion of a large matrix $\mathbf{U}_z$ constructed from all N_0 samples. The least complex methods are histogram-based SSMs whose complexity is the same as for conventional DPD adaptation using a block of continuous feedback samples. However, the conventional methods provide poor linearisation capabilities for a small number of samples N as shown in the rest of the paper. Note that for the same number of selected samples N, QRS and GSS are of higher computational complexity than the conventional DPD. The computational complexity is reduced as QRS and GSS require lower N.

Although we have neglected memory requirements in our analysis, please note that QRS and GSS require much more memory to store the whole $N_0\times P$ matrix $\mathbf{U}_z$ and the $N_0\times 1$ vector Δ than histogram-based SSMs, which only need to store the reduced $N\times P$ matrix $\mathbf{U}_z$ and the $N\times 1$ vector Δ.

E. Reduction of Hardware Resources

Since SSM allows undersampling the PA output, it allows for ADCs with decreased sampling frequency in the feedback. However, the ADC analogue input bandwidth and the speed of ADC sample-and-hold circuit have to be sufficient to cover the desired signal bandwidth plus close intermodulation products. In common state of the art, it is required that the feedback path bandwidth be three to five times the desired channel bandwidth [33], [34]. ADCs with high analogue input bandwidth are commonly employed in current communications systems either for Sub-Nyquist sampling or in interleaved ADCs.

The main advantages of the decreased sampling frequency of the feedback ADCs are the lower power consumption, the decreased system complexity, and the price. We present these parameters on the example transmitter from Fig. 1. We consider two cases: in one case the conventional DPD and in the other case an undersampled ADC in the feedback. The communication signal bandwidth was chosen to be 300 MHz,

TABLE II
COMPARISON OF SYSTEM PARAMETERS

	Conventional DPD 2× AD9690	DPD with SSM 2× AD9629
F_S	1000 MSps	20 MSps
Max. B_{in}	500 MHz	700 MHz
ENOB	10.5 bits	11 bits
Digital interface	JESD204B	Parallel
Power consumption	≈4 W	≈90 mW
Price	≈ 720 USD	≈ 20 USD

which implies a required feedback bandwidth of at least 900 MHz. The conventional DPD requires two ADCs with the sampling frequency $F_S \geq$ 900MSps. DPD with SSM requires two ADCs for IQ sampling with the input bandwidth $B_{in} \geq$ 450 MHz and arbitrary F_S. For this example we have selected ADCs AD9690 with $F_S = 1$ GSps for the conventional DPD and AD9629 with $B_{in} = 700$ MHz and $F_S = 20$ MSps for DPD with SSM, both labelled as low power ADCs. Table II gives a comparison of a transmitter with the conventional DPD and a transmitter with the undersampled feedback allowed by the proposed SSM. The values were taken from ADC data sheets [35], [36]. The comparison of system parameters in Table II does not include the power consumption and price of auxiliary components, e.g., clock generators, buffers, filters, etc.

V. SIMULATIONS

We simulate the DPD performance with the proposed SSM described in Sec. III. Additionally, all simulations include the conventional DPD without SSM as a reference. For all methods, DLA has been used. Since it is iterative, each simulation run consists of $M = 80$ iterations. At the beginning of the m-th iteration the transmit signal $\mathbf{z}$ is generated with random data symbols and used for all simulated SSMs. Afterwards in the m-th iteration, the following steps bound to the specific SSM are executed:

1) signal $\mathbf{z}$ is predistorted using coefficients $\mathbf{b}'(m)$ to obtain the PA input $\mathbf{x}$,
2) signal $\mathbf{x}$ is sent through the PA model to get the PA output $\mathbf{y}$,
3) new DPD coefficients $\mathbf{b}'(m+1)$ are calculated,
4) the linearisation performance metrics are evaluated.

For the first iteration, the DPD coefficients $\mathbf{b}'(0)$ are set to $[0.5, 0, \ldots, 0]^T$. The step size parameter μ was set to 0.1 as a decent trade-off between the convergence probability and the convergence speed. The whole process is repeated 1000 times to evaluate the 95% confidence intervals of the results.

A. PA Model for Simulations

In simulations we employ the PA model based on GMP [28]. We have chosen GMP, because it can be sufficiently complex to accurately model the state-of-the-art PAs. Moreover, for realistic simulation results it is necessary to apply the PA model of higher complexity than the DPD model. GMP is defined as [28]

$$\begin{aligned} y(n) = &\sum_{k\in K_a}\sum_{l\in L_a} b_{0,k,l}\,x(n-l)|x(n-l)|^k \\ &+ \sum_{k\in K_b}\sum_{l\in L_b}\sum_{m\in M_b} b_{1,k,l,m}\,x(n-l)|x(n-l-m)|^k \\ &+ \sum_{k\in K_c}\sum_{l\in L_c}\sum_{m\in M_c} b_{2,k,l,m}\,x(n-l)|x(n-l+m)|^k, \end{aligned} \quad (24)$$

where $b_{0,k,l}$, $b_{1,k,l,m}$, and $b_{2,k,l,m}$ are the PA model coefficients; K_a and L_a are the index arrays for aligned signal and envelope; K_b, L_b and M_b are the index arrays for signal and lagging envelope; and K_c, L_c and M_c are index arrays for signal and leading envelope. We include models of the following PAs in our simulations:

PA1: Two-stage PA in class AB designed for SATCOM applications at 1625 MHz, with monolithic PA GALI-24+ from Mini-Circuits in the first stage and the GaN SiC HEMT transistor TGF2965 from Qorvo in the second stage. The first stage power supply voltage was set to 8 V, the second stage power supply voltage was set to 32 V, and the gate voltage was set to ensure that the drain quiescent current was 20 mA. The total gain of both stages is approximately 30 dB.

PA2: Class-F power amplifier designed for SATCOM applications at 1625 MHz, with output power P_{3dB} = 43 dBm, based on the GaN SiC HEMT transistor T2G6003028 from Qorvo. The PA gain is approximately 12 dB, the power supply voltage 28 V, the gate voltage −3.03 V. Together with this PA, the 5-W amplifier Minicircuits ZHL-5W-2G-S+ was used as a predriver.

PA3: Block amplifier ADL5610.

All PA models were extracted from measurements at a centre frequency of 1600 MHz, excited by a test signal with the bandwidth $B \approx 6$ MHz, further described in Sec. V-C. Index arrays of PA models were set to ensure sufficient fidelity of the models. The particular indexes were set to $K_a = \{0, 1, 2, 3, 4, 5\}$, $L_a = \{0, 1, 2, 3\}$, $K_b = \{2, 4\}$, $L_b = \{0\}$, $M_b = \{1\}$, $K_c = \{2, 4\}$, $L_c = \{0\}$, $M_c = \{1\}$. The PA model coefficients can be found in the source codes provided.

B. DPD Model

As a model of digital predistorter we have chosen the DDR2 model [29], because the simple MP model did not achieve sufficient linearisation results for the tested amplifiers. DDR2 is defined as [29]

$$\begin{aligned} x(n) = &\sum_{k=0}^{\frac{K-1}{2}}\sum_{q=0}^{Q} b'_{0,k,q}|z(n)|^{2k}z(n-q) \\ &+ \sum_{k=1}^{\frac{K-1}{2}}\sum_{q=1}^{Q} b'_{1,k,q}|z(n)|^{2(k-1)}z^2(n)z^*(n-q) \\ &+ \sum_{k=1}^{\frac{K-1}{2}}\sum_{q=1}^{Q} b'_{2,k,q}|z(n)|^{2(k-1)}z(n)|z(n-q)|^2 \\ &+ \sum_{k=1}^{\frac{K-1}{2}}\sum_{q=1}^{Q} b'_{3,k,q}|z(n)|^{2(k-1)}z^*(n)z^2(n-q), \end{aligned} \quad (25)$$

where $b'_{0,k,q}$, $b'_{1,k,q}$, $b'_{2,k,q}$, $b'_{3,k,q}$ are the DPD coefficients. For the following simulations, we have chosen the DPD order to be $K = 7$ and $Q = 1$, which yields $P = 17$ DPD coefficients.

C. Test Signal

We have chosen the filtered orthogonal frequency-division multiplexing (F-OFDM) signal waveform with the 64-state quadrature amplitude modulation (64-QAM) as the internal modulation to demonstrate the linearisation capabilities of the proposed SSM for DPD adaptation. The filter for F-OFDM has been designed as proposed in [37], [38]. The filter coefficients $f_B(u)$ are given as

$$f_B(u) = \frac{p_B(u)\, w(u)}{\sum_n p_B(u)\, w(u)}, \tag{26}$$

where $p_B(u)$ is the sinc function and $w(u)$ is the window function. The sinc function is defined as

$$p_B(u) = \mathrm{sinc}\left((W + 2\,\Delta W)\frac{u}{Z}\right), \tag{27}$$

where Z is the FFT size, W is the number of assigned data subcarriers, and ΔW is the tone-offset, where $2\Delta W$ is the difference between the desired passband and the designed sinc filter passband [38]. The window function is defined as

$$w(u) = \left(0.5\left(1 + \cos\left(\frac{2\pi u}{L-1}\right)\right)\right)^{0.6}, \tag{28}$$

where $-\lfloor \frac{L}{2} \rfloor \leq u \leq \lfloor \frac{L}{2} \rfloor$ and L is the number of filter taps.

In each iteration, we generate the F-OFDM signal with 12 frames, each frame with 68 resource blocks with a block size of 12 subcarriers, resulting in $W = 12 \cdot 68 = 816$. The FFT size is set to $Z = 4096$ and the filter length is $L = 2049$. The sampling frequency is limited by the measurement setup and is set to $F_S = 30\,\mathrm{MHz}$. These parameters provide the test signal vector with the channel bandwidth $B \approx F_S/5 \approx 6\,\mathrm{MHz}$ and $N_0 = 52\,064$ samples. The signal mean power was set constant during the DPD adaptation such that the mean PA output power over the last 20 iterations is equal for all the methods.

D. Linearisation Performance Metrics

The linearisation performance is qualified based on NMSE between the desired signal $\mathbf{z}$ and the actual scaled PA output $\mathbf{y}$, the error vector magnitude (EVM) between the demodulated data symbols $\mathbf{s}_{\mathrm{rx}}$ and transmitted data symbols $\mathbf{s}_{\mathrm{tx}}$, and adjacent channel power ratio (ACPR) in the first adjacent channels. We use following definitions for these metrics

$$\mathrm{NMSE(dB)} = 10\log_{10}\frac{\sum_n |y(n) - z(n)|^2}{\sum_n |z(n)|^2}, \tag{29}$$

$$\mathrm{EVM}(\%) = \frac{\sum_n |s_{\mathrm{rx}}(n) - s_{\mathrm{tx}}(n)|^2}{\sum_n |s_{\mathrm{tx}}(n)|^2} \cdot 100\%, \tag{30}$$

$$\mathrm{ACPR(dB)} = 10\log_{10}\frac{P_{\mathrm{ladj}} + P_{\mathrm{radj}}}{2P_{\mathrm{main}}}, \tag{31}$$

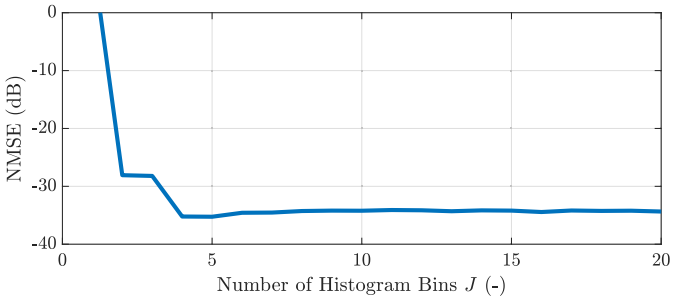

Fig. 5. Simulated influence of the number of histogram bins on achievable NMSE using DPD with EDH and $N = 30$.

where P_{main} is the power in the main channel, and P_{ladj}, P_{radj} are the powers in the left and right adjacent channels respectively. The adjacent channels are of width B and frequency offset $1.1\,B$.

Hereinafter presented metrics are averaged over the last 20 iteration cycles to suppress the influence of different realisations of the transmitted signal.

E. Settings Specific to Histogram-Based Methods

We have set the number J of histogram bins for EDH and GOH based on the simulation presented in Fig. 5, where we have analysed the influence of the number of bins on achievable NMSE, using DPD with EDH. We can see that for DPD adaptation the sufficient number of histogram bins J is 5. We set $J = 10$ to ensure a sufficient margin, as we expect a very low computational resource allocation for histogram-based SSMs.

F. Histogram Optimisation by Genetic Algorithm

We have optimised histogram target bin counts d_j by the genetic algorithm [32] for each simulated number of selected samples N. Optimised bin counts are integers from the interval $[0, N]$ and need to fulfil the condition $\sum d_j = N$. We have set the optimisation parameters as follows: the population size 100, the maximum number of generations 20, the population fraction at the next generation created by crossover 0.8, the probability of mutation 1%. We have defined the objective function as an average of the NMSE results over 10 runs. Each run consists of 80 iterations and the NMSE results for averaging are taken only from the last 20 iterations.

G. Simulation Results

Hereinafter we present detailed simulation results for the model PA1 in Fig. 6, Fig. 7, and Fig. 8. The simulation results for the other models, PA2 and PA3, are provided to verify SSMs in a condensed form in Table III and Table IV.

The relation between NMSE and the number of selected samples N is depicted in Fig. 6. We can observe that the conventional DPD starts to improve the transmitter linearity when more than 1300 samples are selected for DPD adaptation. US follows the behaviour of the conventional DPD but with slightly less needed samples. QRS starts to improve linearity of the output signal from 18 selected samples, but up

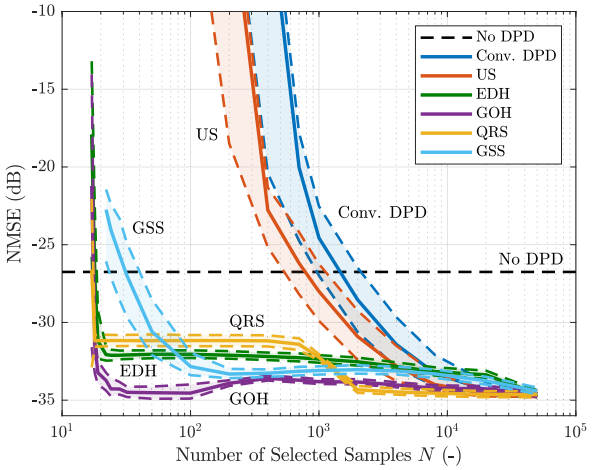

Fig. 6. Simulation results of NMSE as a function of the number of selected samples N with 95% confidence intervals depicted by coloured dashed lines for the model PA1. The black dashed line represents simulated NMSE of the PA output without DPD.

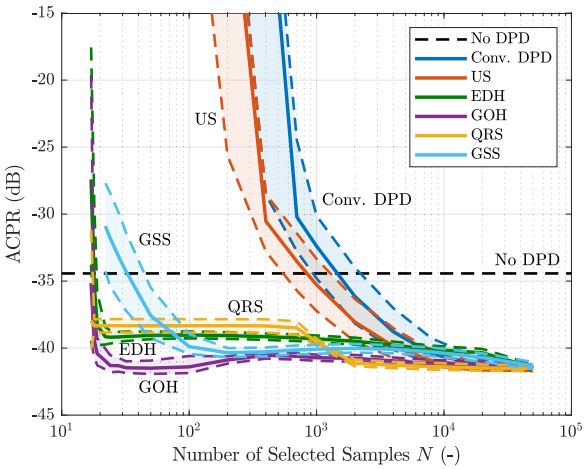

Fig. 8. Simulation results of ACPR as a function of the number of selected samples N with 95% confidence intervals depicted by coloured dashed lines for the model PA1. The black dashed line represents simulated ACPR of the PA output without DPD.

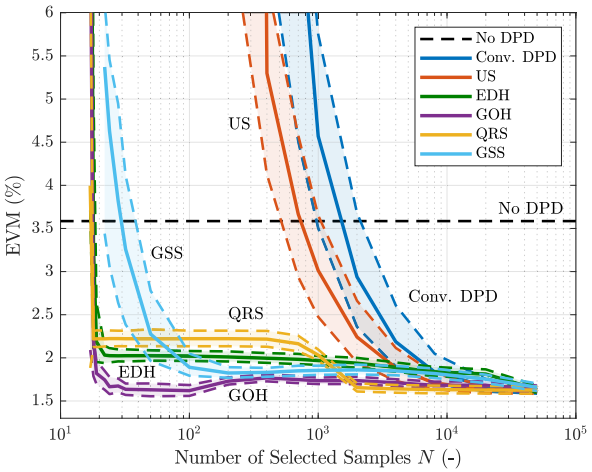

Fig. 7. Simulation results of EVM as a function of the number of selected samples N with 95% confidence intervals depicted by coloured dashed lines for the model PA1. The black dashed line represents simulated EVM of the PA output without DPD. EVM of the generated signal is 1.1% due to the nonorthogonality caused by the inherent F-OFDM filtering.

to 2000 samples it provides suboptimal results. This is mainly due to improving problem conditioning but considering neither the signal statistics nor the observed feedback samples, as has been discussed above. Please recall that we have 17 DPD coefficients and taking only 18 samples is almost equivalent to solving a fully determined system.

GSS shows good linearisation performance for higher than 100 selected samples. We suppose this is caused by choosing samples on the random basis with respect to the probabilities calculated by GS. For a low number of selected samples there is no margin for selecting unimportant samples. EDH starts to linearise from 19 selected samples and slightly improves with increasing number of selected samples. For $N < 100$, EDH outperforms GS and for $N > 100$, EDH provides NMSE less than 0.8 dB higher than GS.

In the region up to $N < 2000$ samples, GOH outperforms all other methods. For $N > 2000$ GOH is very close to QRS. Please note that this difference is very small, about 0.5 dB in terms of NMSE. For GOH, in the region from 100 samples, we can see the effect of a larger space in which the genetic algorithm looks for the optimal histogram. In this region NMSE is up to 1 dB worse than for $N = 100$ or $N = 50\,000$. This effect is caused by the fixed number of maximum generations and by the population size of the genetic algorithm, set irrespective of the number of selected samples N.

EVM as a function of the number of selected samples N is depicted in Fig. 7. The simulated EVM results agree with the NMSE results shown in Fig. 6. The ACPR results are presented in Fig. 8 and similarly agree with the NMSE results. One can notice that for PA1 the linearisation performance in terms of the ACPR improvement is not impressive. This is likely caused by specific characteristics of PA1 and the DPD model. The key point is that already for a low N the proposed methods provide the linearisation performance of the conventional DPD achieved only for high N. A higher ACPR improvement is achieved for PA3 with results in Table IV, as described below.

Summary results for PA2 and PA3 are presented in Table III and Table IV respectively. Although EDH shows good linearisation capabilities for a low number of samples, GOH outperforms other SSMs for a low number of selected samples in all cases and is close to the other methods for a high number of samples.

Model PA2 in Table III is highly nonlinear, as PA2 is designed in class F, and we can see that the conventional DPD and the simple undersampling do not converge for a low number of samples $N < 10\,000$. On the other hand, the model PA3 with results shown in Table IV is less nonlinear and the conventional DPD and US provide a decent linearisation for $N = 1000$ selected samples. In Table IV we can notice that the PA nonlinearity causes mainly leakage into adjacent channels while the signal in the main channel remains undistorted, as illustrated by EVM.

TABLE III
SIMULATION RESULTS FOR MODEL PA2

Metric	N	Conv. DPD	US	EDH	GOH
NMSE (dB)	22	DNC	DNC	-15.4	-20.4
	40	DNC	DNC	-19.8	-20.8
	10 000	-14	-18	-20.8	-21.1
	20 000	-20.6	-21.0	-20.9	-21.1
EVM (%)	22	DNC	DNC	9.8	4.4
	40	DNC	DNC	4.9	4.2
	10 000	13.2	7.2	4.1	3.9
	20 000	4.2	4.0	4.0	3.9
ACPR (dB)	22	DNC	DNC	-22.5	-26.8
	40	DNC	DNC	-26.2	-27.2
	10 000	-21	-26.4	-27.1	-27.2
	20 000	-27	-27.2	-27.2	-27.3

Conventional DPD and US do not converge (DNC) for a low number of selected samples N. Simulated metrics without DPD: NMSE = -13.9 dB, EVM = 13.2%, and ACPR = -21.8 dB. EVM of the generated signal is 1.1% due to the nonorthogonality caused by the inherent F-OFDM filtering.

TABLE IV
SIMULATION RESULTS FOR MODEL PA3

Metric	N	Conv. DPD	US	EDH	GOH
NMSE (dB)	19	DNC	DNC	DNC	-45.8
	22	DNC	DNC	-46.3	-46.7
	1000	-44.8	-46.0	-47.0	-47.0
	4000	-46.7	-46.6	-47.0	-47.1
EVM (%)	19	DNC	DNC	DNC	1.15
	22	DNC	DNC	1.15	1.15
	1000	1.17	1.16	1.15	1.15
	4000	1.95	1.15	1.15	1.15
ACPR (dB)	19	DNC	DNC	DNC	-58.2
	22	DNC	DNC	-60.4	-61.8
	1000	-56.5	-59.5	-64.4	-64.9
	4000	-62.4	-61.8	-64.8	-65.0

Conventional DPD and US do not converge (DNC) for a low number of selected samples N. Simulated metrics without DPD: NMSE = -33.3 dB, EVM = 1.5%, and ACPR = -42.6 dB. EVM of the generated signal is 1.1% due to the nonorthogonality caused by the inherent F-OFDM filtering.

VI. MEASUREMENTS

In addition to simulations for all models PA1, PA2, and PA3, we evaluate the SSMs for PA1 in measurements. Our measurement setup (Fig. 9) employs the vector signal generator Rohde & Schwarz SMU200A to generate the input signal of PA1. The PA1 output is connected to the real time spectrum analyser Rohde&Schwarz FSVR through a high power RF attenuator with an attenuation of 10 dB and a maximum dissipated power of 50 W followed by two smaller RF attenuators of 10 dB each. We use the digital multimeter Keysight 34461A as an Ampere metre to set the drain quiescent current of the PA1 second-stage transistor.

We have performed the measurements with the same settings as for simulations. The only change is that for measurements we perform the 80 iterations only once, and hence the values presented are not averaged.

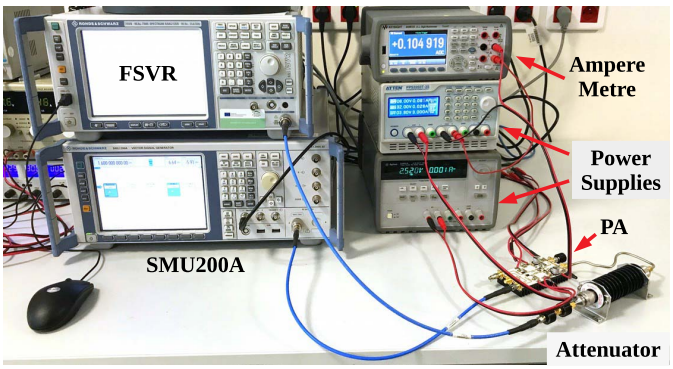

Fig. 9. Measurement setup with PA1, the vector signal generator SMU200A, and the real time spectrum analyser FSVR.

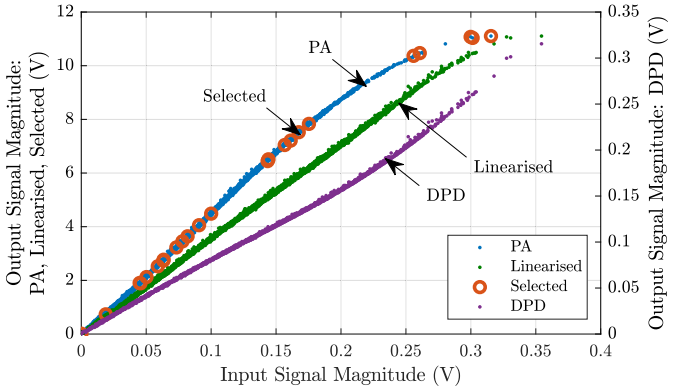

Fig. 10. Measured AM/AM characteristics of PA1, DPD, and the whole linearised transmitter. The depicted $N = 24$ samples were selected by GOH and used for DPD adaptation.

In Fig. 10 we present the measured AM/AM characteristics of PA1, DPD and the whole linearised transmitter. The characteristics of DPD and the linearised transmitter are depicted for DPD adapted using GOH with $N = 24$ selected samples. Please notice that the selected samples are not distributed evenly over the input magnitudes, but rather concentrated in the region with lower magnitudes to cover the region with the high signal occurrence probability and then in the region with higher magnitudes to cover the most nonlinear region of PA.

The measured power spectral density (PSD) of the PA1 output linearised with DPD adapted using different techniques is depicted in Fig. 11. The PSD of the PA1 output without DPD is shown as a reference. The results are shown for $N = 400$ selected samples. Please note that for $N = 400$ the conventional DPD causes higher out-of-band emissions than PA1 without DPD. US slightly improves the out-of-band emissions, while GOH in this case provides the best linearisation.

Measured DPD performance in terms of NMSE, EVM, and ACPR is shown in Fig. 12, Fig. 13, Fig. 14 respectively. We can see that the conventional DPD and US in real measurements work slightly better than in simulations. This could signify that the extracted PA model is more difficult to linearise by these methods than the real PA. The trend for these methods follows the simulation results, and we can see

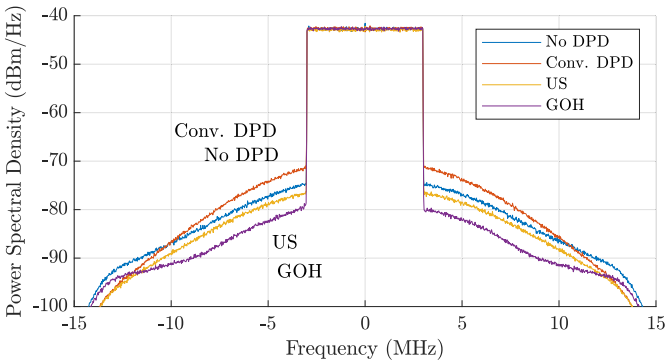

Fig. 11. Measured power spectral density of PA1 output without DPD and with DPD adapted by different SSMs for the number of selected samples $N = 400$.

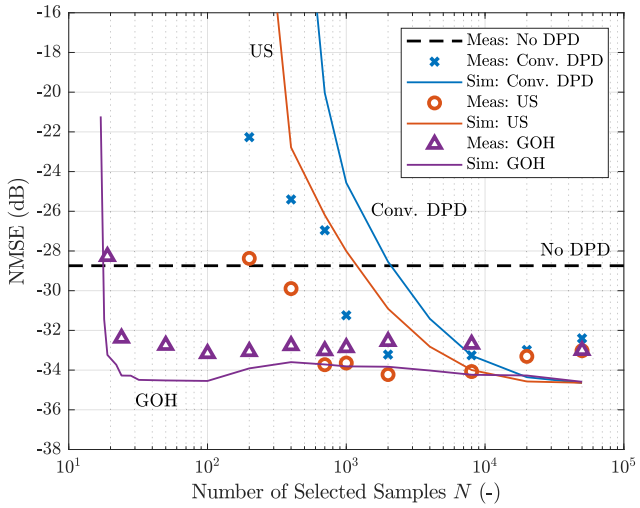

Fig. 12. Measurement results of NMSE as a function of the number of selected samples N compared to the simulation results. The black dashed line represents the measured NMSE of the PA1 output without DPD.

a spread of measured points from this trend. This spread is in compliance with the wider 95% confidence intervals shown in the simulation results. The measured NMSE for GOH is around 1 dB worse with respect to the average NMSE in the simulations. This increase could be probably caused by the measurement noise which is neglected in simulations. Measurement results in terms of EVM and ACPR follow the NMSE measurement results which exactly complies with the simulations.

A. High-Bandwidth Measurements

We have verified the proposed methods also by measurements with a high-bandwidth signal. We have performed this measurement on our 60-GHz DPD setup which is based on Infineon BGT-60 and has sampling frequency $F_S = 1\,\text{GSps}$, for more details see [39]. We have applied the test signal with bandwidth $B = 100\,\text{MHz}$.

The measured PSD of the setup output without DPD and with DPD adapted by different SSMs is shown in Fig. 15. The PSD is measured for the number of selected samples $N = 200$. We can see that GOH outperforms the conventional DPD and

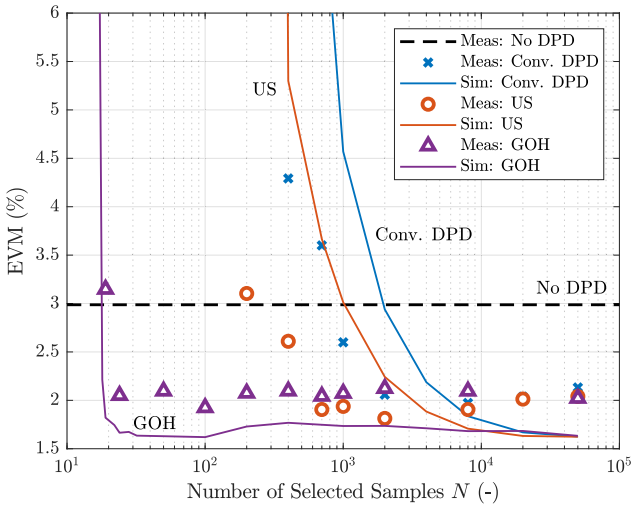

Fig. 13. Measurement results of EVM as a function of the number of selected samples N compared to the simulation results. The black dashed line represents the measured EVM of the PA1 output without DPD. EVM of the generated signal is 1.1% due to the nonorthogonality caused by the inherent F-OFDM filtering.

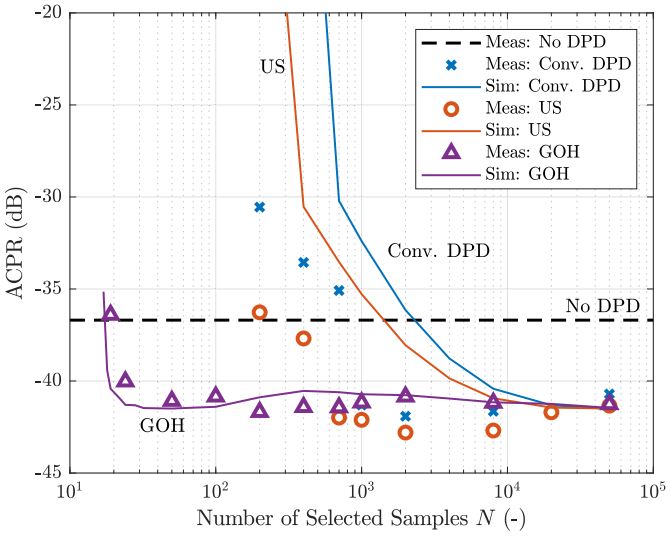

Fig. 14. Measurement results of ACPR as a function of the number of selected samples N compared to the simulation results. The black dashed line represents the simulated ACPR of the PA1 output without DPD.

TABLE V
MEASUREMENT RESULTS FOR 60- GHz SETUP

	No DPD	Conv. DPD	US	GOH
NMSE (dB)	-24.5	-22.0	-29.6	-32.0
EVM (%)	5.8	6.8	3.3	2.7
ACPR (dB)	-39.1	-31.7	-43.4	-49.0

also US. NMSE, EVM and ACPR for this measurement are stated in Table V. We would like to remark that during the measurement we had substantial issues with the stability of the conventional DPD adaptation and we have observed the similar tendency also for US. On the contrary, we have not recognised any instability issues caused by GOH.

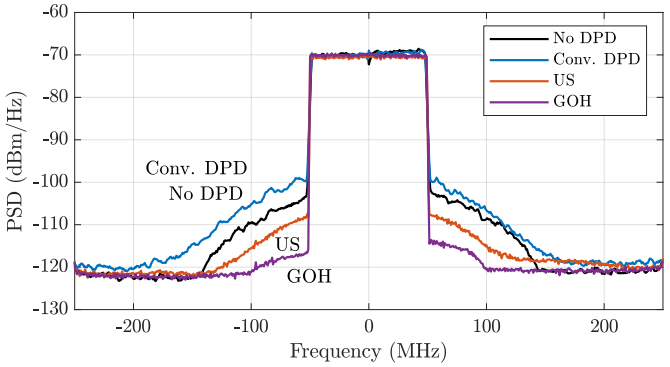

Fig. 15. The measured power spectral density at the 60-GHz setup output without DPD and with DPD adapted by different SSMs for the number of selected samples $N = 200$.

VII. Conclusion

The number of samples for DPD adaptation directly influences the computational complexity of DPD adaptation. In this paper we have proposed sample selection methods for DPD adaptation with the intention to minimise the required number of samples for DPD adaptation and thus to minimise the computational complexity of DPD adaptation. We have shown that the proposed GOH outperforms other sample selection methods in terms of linearisation capabilities. For a very low number of selected samples, GOH provides a linearisation performance equivalent to the maximum achievable linearisation performance of the conventional DPD. Simulations on the model of a two-stage power amplifier (PA1) designed for SATCOM applications have revealed that GOH can achieve a sufficient linearisation performance already for $N = 24$ selected samples while the conventional DPD achieves the equivalent performance only for $N \geq 10^4$. Since the computational complexity is linear with respect to the number of required samples, this indicates a 400-times improvement over the conventional DPD in terms of computational complexity.

Acknowledgment

The authors would like to thank our colleague M. Pospisil for helping us with high-bandwidth measurements on our 60-GHz setup and for his effort making this setup functional. They would like to express appreciation to our colleagues from TU Wien, especially to Erich Zöchmann and Stefan Schwarz, for their scientific advices and suggestions.

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] J. Chani-Cahuana, M. Ozen, C. Fager, and T. Eriksson, "Digital predistortion parameter identification for RF power amplifiers using real-valued output data," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 64, no. 10, pp. 1227–1231, Oct. 2017.

[3] J. Kral, T. Gothans, R. Marsalek, and M. Harvanek, "Digital predistorter with real-valued feedback employing forward model estimation," in *Proc. 25th Int. Conf. Telecommun. (ICT)*, Jun. 2018, pp. 471–475.

[4] Y. Liu, J. J. Yan, H.-T. Dabag, and P. M. Asbeck, "Novel technique for wideband digital predistortion of power amplifiers with an under-sampling ADC," *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 11, pp. 2604–2617, Nov. 2014.

[5] L. Zhang and Y. Feng, "An improved digital predistortion in wideband wireless transmitters using an under-sampled feedback loop," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 910–913, May 2016.

[6] Z. Wang, S. Ibrahim, H. Su, and R. Farrell, "Generalised digital predistortion of RF power amplifiers with low-rate feedback signal," in *Proc. 46th Eur. Microw. Conf. (EuMC)*, Oct. 2016, pp. 831–834.

[7] Z. Wang, W. Chen, G. Su, F. M. Ghannouchi, Z. Feng, and Y. Liu, "Low feedback sampling rate digital predistortion for wideband wireless transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 11, pp. 3528–3539, Nov. 2016.

[8] H. Paaso and A. Mammela, "Comparison of direct learning and indirect learning predistortion architectures," in *Proc. IEEE Int. Symp. Wireless Commun. Syst.*, Oct. 2008, pp. 309–313.

[9] H. Huang, P. Mitran, and S. Boumaiza, "Digital predistortion function synthesis using undersampled feedback signal," *IEEE Microw. Wireless Compon. Lett.*, vol. 26, no. 10, pp. 855–857, Oct. 2016.

[10] Z. Wang, L. Guan, and R. Farrell, "Compact undersampled digital predistortion for flexible single-chain multi-band RF transmitter," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2017, pp. 1542–1545.

[11] 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.

[12] Q. Zhang, W. Chen, Z. Wang, and G. Su, "A single feedback architecture for dual-band digital predistortion with under-sampling technique," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Mar. 2016, pp. 1–4.

[13] N. Guan, N. Wu, and H. Wang, "Digital predistortion of wideband power amplifier with single undersampling ADC," *IEEE Microw. Wireless Compon. Lett.*, vol. 27, no. 11, pp. 1016–1018, Nov. 2017.

[14] N. Hammler, A. Cathelin, P. Cathelin, and B. Murmann, "A spectrum-sensing DPD feedback receiver with 30× reduction in ADC acquisition bandwidth and sample rate," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 66, no. 9, pp. 3340–3351, Sep. 2019.

[15] L. Guan and A. Zhu, "Optimized low-complexity implementation of least squares based model extraction for digital predistortion of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 3, pp. 594–603, Mar. 2012.

[16] N. Kelly and A. Zhu, "Low-complexity stochastic optimization-based model extraction for digital predistortion of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 5, pp. 1373–1382, May 2016.

[17] Z. Wang, W. Chen, G. Su, F. M. Ghannouchi, Z. Feng, and Y. Liu, "Low computational complexity digital predistortion based on direct learning with covariance matrix," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 11, pp. 4274–4284, Nov. 2017.

[18] R. B. Benedict and W. Honcharenko, "Selecting samples for amplifier digital predistortion estimation," U.S. Patent 7 466 197, Dec. 16, 2008.

[19] R. B. Benedict, "Selecting samples for amplifier digital predistortion estimation," U.S. Patent 7 479 828, Jan. 20, 2009.

[20] D. Huang, X. Huang, and H. Leung, "Nonlinear compensation of high power amplifier distortion for communication using a histogram-based method," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4343–4351, Nov. 2006.

[21] T. Wang, P. L. Gilabert, and G. Montoro, "Under-sampling effects and computational cost reduction in RF power amplifier behavioral modeling," in *Proc. 10th Eur. Microw. Integr. Circuits Conf. (EuMIC)*, Sep. 2015, pp. 57–60.

[22] Z. Wang, J. Dooley, K. Finnerty, and R. Farrell, "Selection of compressed training data for RF power amplifier behavioral modeling," in *Proc. 10th Eur. Microw. Integr. Circuits Conf. (EuMIC)*, Sep. 2015, pp. 53–56.

[23] T. F. Chan, "Rank revealing QR-factorizations," *Linear Algebra Appl.*, vols. 88–89, pp. 67–82, Apr. 1987. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0024379587901030

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

[25] D. R. Morgan, Z. Ma, and L. Ding, "Reducing measurement noise effects in digital predistortion of RF power amplifiers," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 4, May 2003, pp. 2436–2439.

[26] P. N. Landin, A. E. Mayer, and T. Eriksson, "MILA—A noise mitigation technique for RF power amplifier linearization," in *Proc. IEEE 11th Int. Multi-Conference Syst., Signals Devices (SSD)*, Feb. 2014, pp. 1–4.

[27] J. Kim and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory," *Electron. Lett.*, vol. 37, no. 23, pp. 1417–1418, 2001.
[28] 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.
[29] L. Guan and A. Zhu, "Simplified dynamic deviation reduction-based volterra model for doherty power amplifiers," in *Proc. Workshop Integr. Nonlinear Microw. Millimetre-Wave Circuits*, Apr. 2011, pp. 1–4.
[30] Z. Yu, H. Xie, and E. Zhu, "Wide-band linear characteristic compensation for DPD systems with direct learning," in *Proc. Asia–Pacific Microw. Conf. Proc. (APMC)*, Nov. 2013, pp. 790–793.
[31] A. A. Rontogiannis and S. Theodoridis, "New fast inverse QR least squares adaptive algorithms," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, vol. 2, May 1995, pp. 1412–1415.
[32] O. Kramer, *Genetic Algorithm Essentials* (Studies in Computational Intelligence), vol. 679. Oldenburg, Germany: Springer, 2017.
[33] F. M. Ghannouchi and O. Hammi, "Behavioral modeling and predistortion," *IEEE Microw. Mag.*, vol. 10, no. 7, pp. 52–64, Dec. 2009.
[34] J. Zhao, Y. Liu, C. Yu, J. Yu, and S. Li, "A modified band-limited digital predistortion technique for broadband power amplifiers," *IEEE Commun. Lett.*, vol. 20, no. 9, pp. 1800–1803, Sep. 2016.
[35] *Data Sheet AD9690, Rev. B*, Analog Devices, Norwood, MA, USA, May 2017.
[36] *Data Sheet AD9629, Rev. B*, Analog Devices, Norwood, MA, USA, Feb. 2017.
[37] J. Abdoli, M. Jia, and J. Ma, "Filtered OFDM: A new waveform for future wireless systems," in *Proc. IEEE 16th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2015, pp. 66–70.
[38] *f-OFDM Scheme and Filter Design*, document R1-165425, 3GPP, TSG RAN WG1 Meeting, Huawei, HiSilicon, Nanjing, China, May 2016.
[39] R. Marsalek, M. Pospisil, T. Gotthans, and T. Urbanec, "Digital calibration of 60 GHz setup for use in power amplifier predistortion," in *Proc. IEEE 20th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2019, pp. 1–4.

Jan Kral received the bachelor's and Ing. degrees (M.Sc. equivalent) in electrical engineering from the Brno University of Technology, Brno, Czech Republic, in 2012 and 2014, respectively, where he is currently pursuing the Ph.D. degree.

From 2013 to 2014, he worked on his diploma thesis, Fast Beam Current Change Monitor for LHC, CERN, Geneva, Switzerland. His research interests include digital predistortion of power amplifiers, transceiver impairments, and application of digital signal processing to FPGAs.

Tomas Gotthans (Member, IEEE) received the bachelor's and Ing. degrees in electrical engineering from the Brno University of Technology in 2008 and 2010, respectively, and the Ph.D. degree from the Université de Marne-La-Vallée, France, in January 2014. In 2011, he joined ESIEE Paris, ESYCOM Laboratory, where he worked on the project AMBRUN (in collaboration with Thales Communications and Security, TeamCast, Supélec). He is currently an Associate Professor with the Department of Radio Electronics, Brno University of Technology. His research interests include digital predistortion of power amplifiers, wireless communications theory, and chaos theory.

Roman Marsalek (Member, IEEE) received the Ing. degree from the Brno University of Technology, Brno, Czech Republic, in 1999, and the Ph.D. degree from the Université de Marne-La-Vallée, France, in 2003. In 2013, he was a Teaching and a Research Fellow with Johannes Kepler University, Linz, Austria. He is currently a Full Professor with the Department of Radio Electronics, Brno University of Technology. His research interests include wireless communications theory and applied digital signal processing.

Michal Harvanek received the bachelor's and Ing. degrees (M.Sc. equivalent) in electrical engineering from the Brno University of Technology, Brno, Czech Republic, in 2013 and 2015, respectively, where he is currently pursuing the Ph.D. degree with the Department of Radio Electronics. He has collaborated with CISC Semiconductor in the area of RFID signal processing implementations. His research interests include FDM modulation techniques, transceiver impairments, and interference cancellation methods.

Markus Rupp (Fellow, IEEE) received the Dipl.-Ing. degree from the University of Saarbrücken, Saarbrücken, Germany, in 1988, and the Dr.-Ing. degree from Technische Universität Darmstadt, Darmstadt, Germany, in 1993. Until 1995, he had a postdoctoral position at the University of California, Santa Barbara, CA, USA. From 1995 to 2001, he was with the Wireless Technology Research Department, Bell-Labs, Holmdel, NJ, USA. Since October 2001, he has been a Full Professor of digital signal processing in mobile communications with TU Wien, Vienna, Austria.