Inverse Covariance Matrix Estimation for Low-Complexity Closed-Loop DPD Systems: Methods and Performance

Pablo Pascual Campo, *Graduate Student Member, IEEE*, Lauri Anttila, Vesa Lampu, Markus Allén, Yan Guo, Neng Wang, and Mikko Valkama, *Senior Member, IEEE*

Abstract—In this article, we study closed-loop digital predistortion (DPD) systems and associated learning algorithms. Specifically, we propose various low-complexity approaches to estimate and manipulate the inverse of the input data covariance matrix (CM) and combine them with the so-called self-orthogonalized (SO) learning rule. The inherent simplicity of the SO algorithm, combined with the proposed solutions, allows for remarkably reduced complexity in the DPD system while maintaining similar linearization performance compared to other state-of-the-art methods. This is demonstrated with thorough over-the-air (OTA) mmW measurement results at 28 GHz, incorporating a state-of-the-art 64-element active antenna array, and very wide channel bandwidths up to 800 MHz. In addition, complexity analyses are carried out, which together with the measured linearization performance demonstrates favorable performance–complexity tradeoffs in linearizing mmW active array transmitters through the proposed solutions. The techniques can find application in systems where the power amplifier (PA) nonlinearities are time-varying and thus frequent or even constant updating of the DPD is required, good examples being mmW adaptive antenna arrays as well as terminal transmitters in 5G and beyond networks.

Index Terms—Array transmitters, autocorrelation function, Bussgang theorem, closed-loop systems, covariance matrix (CM), digital predistortion, Gauss–Newton (GN), parameter learning, power amplifier (PA), self-orthogonalization.

I. INTRODUCTION

CONTEMPORARY radio communication systems, such as the recently introduced 5G new radio (NR) mobile networks, build on multicarrier modulation—most notably orthogonal frequency-division multiplexing (OFDM). Multicarrier modulations are known to provide spectrally efficient waveforms but also possess high peak-to-average power ratio (PAPR) [1], [2], which complicates the operation of power amplifiers (PAs) close to saturation. In order to control the quality of the transmit waveform while ensuring high power efficiency in the transmit chain, digital predistortion (DPD) is a well-known technical approach—see, e.g., [1], [3]–[5], and the references therein. DPD aims to suppress the unwanted out-of-band (OoB) emissions and in-band nonlinear distortion by applying specific nonlinear preprocessing to the digital transmit waveform. Especially in combination with efficient PAPR mitigation methods [6], DPD can significantly improve the transmitter power efficiency while maintaining at the same time low levels of unwanted nonlinear distortions.

A particularly timely DPD application is the linearization of active antenna arrays, being an essential technology for instance in the emerging 5G NR base stations and mobile devices, especially at the millimeter-wave (mmW) bands. These operating frequencies are also referred to as the so-called frequency range 2 (FR-2) [7] in the 3GPP terminology. Timely examples of articles focusing on this topic include, e.g., [8]–[13]. In such systems, the nonlinear distortion is known to be beam-dependent, stemming from the load-modulation phenomenon [8]. Fast DPD adaptation is thus required such that the nonlinear distortions can be suppressed, while the beam is steered. This issue, along with the extremely high processing rates and channel bandwidths at mmW frequencies, calls for reduced complexity DPD approaches and associated parameter learning algorithms.

Another relatively new DPD use case is mobile device linearization. In general, mobile device power efficiency is very important, which is why the PA is typically operated close to saturation. Thus, despite the more relaxed linearity requirements compared to base-station transmitters, linearization can be pursued for maximum power efficiency. Furthermore, in mobile devices, the transmitted waveforms are very dynamic, as the resource block allocation and output power can change on a per-slot basis [14]. On the other hand, the computational resources in mobile devices are much more limited than in base stations. Thus overall, both the adaptability and the complexity of the DPD system are becoming increasingly important in mobile devices, forming one main motivation for this work and the methods described in this article.

A wide range of DPD models can be found in the existing literature, with some of the most common approaches

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

2

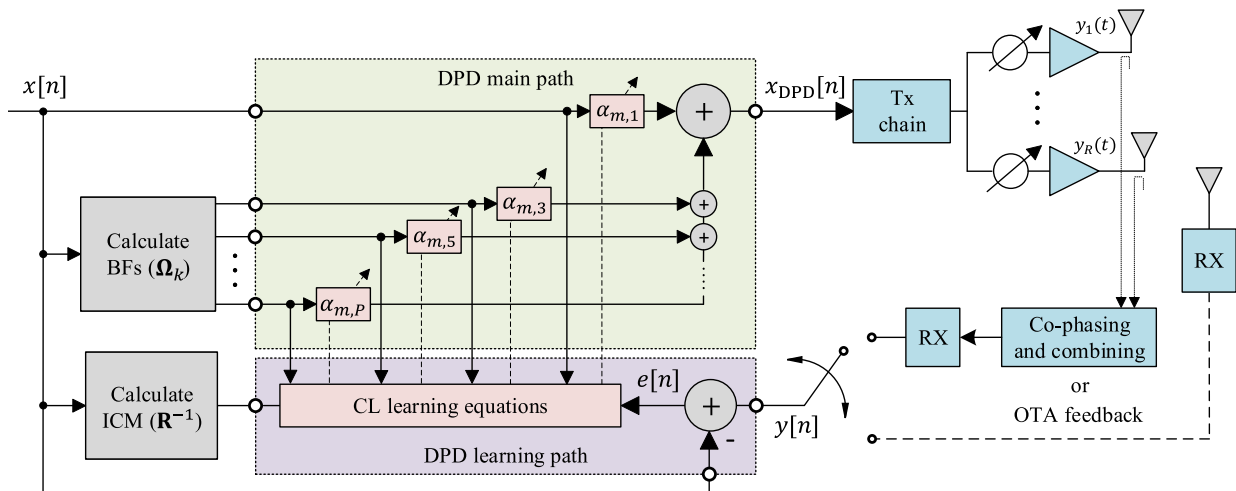IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES

Fig. 1. Illustration of the considered closed-loop MP DPD scheme, with special emphasis on mmW active array linearization with R antenna elements and PA units. The learning signal can be obtained by either co-phasing and combining the individual PA output signals or, alternatively, by using an OTA ORX.

being the memory polynomial (MP) [1], [5], [15] and the generalized MP (GMP) [3], [15], [16]. These approaches can be seen as subsets of the Volterra series [17], [18]. In such methods, the DPD coefficients are usually estimated by using block-based regression algorithms, such as least squares (LS) or Gauss–Newton (GN) [4], [19]. These learning solutions provide reliable coefficient estimation, but they also involve relatively high computational complexity—especially if on-chip estimation were pursued. Alternatively, adaptive filtering algorithms can be used, such as least mean squares (LMS) or some of its variants [20]. The input data basis functions (BFs), however, need to be typically prewhitened or orthogonalized before the actual processing since the elements of the input data are highly correlated, leading to large eigenvalue spread of the input signal covariance matrix (CM) [21].

Literature targeting explicitly low-complexity DPD approaches and associated learning methods is, in turn, somewhat less common. Techniques aiming at this direction are, for instance, [21]–[27]. In [22], a closed-loop MP DPD model was presented, where the model coefficients were estimated with damped GN in combination with a signed regressor algorithm (SRA). However, the signed CM in the SRA suffered from rank-deficiencies, and additional Walsh-Hadamard matrix transformations were required, further increasing the computational complexity. In [21] and [23], a block-adaptive LMS-type algorithm was proposed. However, additional BF orthogonalization or prewhitening had to be applied to ensure fast and reliable convergence. In [27], an SO solution was applied and demonstrated to achieve a similar performance as other state-of-the-art techniques. However, the CM was assumed fixed, and the computational cost of its recalculation was high. In [24]–[26] and [28], cascaded Hammerstein and Wiener structures were proposed. Such approaches typically have less free parameters, leading to solutions with lower associated complexity. However, they are best applied in scenarios where physical knowledge of the system under study is available, which is not always the case. Finally, works in [29] and [30] presented lookup table

(LUT)-based approaches. The LUTs aimed at substituting the high-order polynomials in the DPD model, relaxing the overall complexity. Nonetheless, their modeling capabilities are somewhat limited, depending on the size of the LUTs. To enhance the performance, the LUT sizes can be increased to better describe the nonlinear characteristics, but this leads to higher memory requirements and slower convergence speeds.

In this article, we adopt a closed-loop MP structure in combination with a modified version of the SO learning to update the DPD model coefficients. We adopt the MP structure, in contrast to other more complex solutions, due to the relaxed linearity requirements of the mmW NR FR-2 systems [7], [14] and also because its reduced complexity makes it appealing for various DPD applications, such as the terminal PA linearization. The overall DPD structure is shown in Fig. 1, in the context of mmW active arrays.

The main technical contributions of this article are the development of several low-complexity solutions for estimating and manipulating the inverse CM (ICM) needed in SO learning algorithms. The ICM gathers statistical information of the DPD input data and the corresponding BF samples, and it is an important ingredient in the learning path to ensure good linearization performance and convergence. This is the most computationally heavy element in the SO learning rule and needs to be recalculated whenever the transmit signal type or its spectral characteristics are modified, to ensure optimal DPD execution. Specifically, we propose the following four methods:

1) method to efficiently modify the ICM by removing rows and columns for a smaller DPD parameterization;
2) method to efficiently modify the ICM when the input signal is frequency-shifted, reflecting, e.g., a change in resource allocation inside the channel bandwidth;
3) method to estimate the ICM from the autocorrelation function of the input data, assuming that it is a complex proper Gaussian distributed random signal.

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

PASCUAL CAMPO *et al.*: INVERSE CM ESTIMATION FOR LOW-COMPLEXITY CLOSED-LOOP DPD SYSTEMS 3

4) Method to efficiently approximate the technique in 3) by utilizing the Bussgang decomposition.

Techniques 1) and 2) can be applied to modify a previously calculated ICM with low computational effort, while techniques 3) and 4) require only an estimate of the autocorrelation function of the input data to build the ICM. The SO learning algorithm in combination with the proposed methods allows for fast DPD coefficient estimation with minimal complexity in time-varying scenarios. Hence, on-chip implementations and continuous tracking of the DPD can become feasible.

Extensive over-the-air (OTA) RF measurement results at the 28 GHz mmW center frequency–the 5G NR band n257 [7]—are carried out and reported in order to test and validate the proposed solutions. All measurements feature a state-of-the-art 64-element active antenna array and 5G NR compliant OFDM waveforms, with signal bandwidths ranging from 200 to 400 MHz. In addition, further measurements with an aggregated signal bandwidth of 800 MHz are also conducted, with the aim of pushing the performance boundaries of the considered DPD system. Altogether, the obtained results, alongside with the complexity analyses, indicate very favorable performance–complexity tradeoffs of the proposed solutions when comparing against the state of the art. Finally, we note that although the experiments in this article consider mmW active array linearization, the developed solutions are applicable also in any more classical single-input–single-output DPD system and PA linearization context.

The rest of this article is organized as follows. Section II presents the adopted MP structure and also clarifies the assumptions taken for the mmW setup. Sections III and IV present the proposed low-complexity methods to estimate the ICM, needed for the SO learning rule. Complexity analyses and comparisons are then provided in Section V. Section VI presents the deployed 28 GHz measurement setup and the corresponding FR-2 RF measurement results and their analyses. Finally, conclusions are drawn in Section VII.

Notation Used in This Article: In this article, matrices are represented by capital boldface letters, e.g., $\boldsymbol{\Omega} \in \mathbb{C}^{M \times N}$. Ordinary transpose, Hermitian transpose, and complex conjugation are denoted by $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^*$, respectively. By default, vectors are complex-valued column vectors, presented with lowercase boldface letters, i.e., $\boldsymbol{\alpha} \in \mathbb{C}^{M \times 1} = [\alpha_1 \ \alpha_2 \ \cdots \ \alpha_M]^T$. In addition, the expected value, absolute value, floor, ceil, factorial, Hadamard product, and Kronecker product operators are written as $E\{\cdot\}$, $|\cdot|$, $\lfloor * \rfloor \cdot$, $\lceil * \rceil \cdot$, $!$, $\circ$, and $\otimes$, respectively.

II. CLOSED-LOOP DPD SYSTEM

In this section, the MP DPD model utilized in this work is described. Both the processing and learning paths are detailed, where the actual digital predistortion and the DPD coefficient estimation are executed. The MP model is adopted because it is a widely used model and is known to strike a good balance between linearization performance and computational complexity [1], [3], [5], [21], [23]. In addition, the DPD system builds on a closed-loop structure, where the DPD coefficients are estimated from the input signal, $x[n]$, and the observed learning signal, $y[n]$, [31] according to the notation followed in Fig. 1.

A. DPD Main Path

The complete closed-loop MP DPD scheme is presented in Fig. 1. Here, $x[n]$ is the original baseband signal, $x_{\text{DPD}}[n]$ is the signal after predistortion, and $y_1(t), y_2(t), \ldots, y_R(t)$ are the individual PA output signals to be transmitted. Following this notation and adopting the classical MP model structure, the input–output relation of the digital predistorter can be written as

$$x_{\text{DPD}}[n] = \sum_{\substack{m=0}}^{M-1} \sum_{\substack{p=0 \\ p \text{ odd}}}^{P} \alpha_{m,p} x[n-m] |x[n-m]|^{(p-1)} \quad (1)$$

where M denotes the number of memory taps considered in the model, P corresponds to the polynomial order, and $\alpha_{m,p}$ is the corresponding PA model coefficient for a specific memory tap and polynomial order. Following this configuration, the model has $C = \lceil (P/2) \rceil M$ coefficients. It is noted that while only odd orders are considered in the above model, there are works [32], [33], which have shown that the use of both odd and even orders can have certain benefits, such as the ability to reduce the overall polynomial order P for given linearization performance targets. The ICM estimation methods derived and described in this article focus on MP systems with odd orders, while the extension to cover also even-order polynomials is an interesting topic for future work in this area.

Next, in order to facilitate the closed-loop processing, the input–output relation obtained in (1) can be alternatively expressed with matrix notation as the complex multiplication of the MP BF matrix and the DPD coefficient vector. This expression reads

$$\mathbf{x}_{\text{DPD}} = \boldsymbol{\Omega} \boldsymbol{\alpha} \quad (2)$$

where $\boldsymbol{\alpha} \in \mathbb{C}^{C \times 1}$ contains the PA model coefficients and $\boldsymbol{\Omega} \in \mathbb{C}^{N \times C}$ is the original monomial BF matrix with N denoting the total number of the available input data samples. $\boldsymbol{\Omega}$ can be further expressed as

$$\boldsymbol{\Omega} = \begin{bmatrix} \mathbf{x}_{1,0} \ \mathbf{x}_{1,1} \ \cdots \ \mathbf{x}_{1,M-1} \ \mathbf{x}_{3,0} \ \mathbf{x}_{3,1} \ \cdots \ \mathbf{x}_{3,M-1} \ \cdots \\ \cdots \ \mathbf{x}_{P,0} \ \mathbf{x}_{P,1} \ \cdots \ \mathbf{x}_{P,M-1} \end{bmatrix} \quad (3)$$

with each vector $\mathbf{x}_{p,m} \in \mathbb{C}^{N \times 1}$ being defined as

$$\mathbf{x}_{p,m} = \begin{bmatrix} x[n-m]|x[n-m]|^{p-1} \\ x[n-m+1]|x[n-m+1]|^{p-1} \\ \vdots \\ x[n-m+N-1]|x[n-m+N-1]|^{p-1} \end{bmatrix}. \quad (4)$$

The order of the BF terms in (3) becomes important for the formulation taken in the methods presented in Sections III and IV. We finally note that in the first DPD iteration, the vector $\boldsymbol{\alpha}$ is usually initialized with a one in the first element and zeros in the rest (i.e., $\boldsymbol{\alpha} = [1 \ 0 \cdots 0]^T$) such that only the unmodified linear term is passed through the predistorter.

B. DPD Learning Path

In closed-loop DPD systems, the observation signal $y[n]$ is nonlinear in the DPD parameters. Therefore, nonlinear LS

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

4

IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES

techniques need to be applied, such as those based on the GN algorithm [29], [34]. At block iteration k, these iterative algorithms build on the error vector $\mathbf{e}_k = \mathbf{y}_k - \mathbf{x}_k$ that is stacking the prevailing error signal samples of the current processing block of size K. In this article, we focus on the SO and the damped GN learning methods [21], [22], [29]. In Sections III and IV, the SO learning rule is studied in detail and combined with different methods to reduce the involved computational complexity, while the GN serves as a reference solution. To this end, the SO and the GN learning rules read

$$\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k + \mu_s (\mathbf{R}^*)^{-1} \boldsymbol{\Omega}_k^H \mathbf{e}_k \tag{5}$$

$$\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k + \mu_g (\boldsymbol{\Omega}_k^H \boldsymbol{\Omega}_k)^{-1} \boldsymbol{\Omega}_k^H \mathbf{e}_k \tag{6}$$

respectively, where μ_s and μ_g are the learning steps for the updates, $\boldsymbol{\Omega}_k \in \mathbb{C}^{K \times C}$ is the BF matrix corresponding to the DPD block iteration k, and $\mathbf{R} \in \mathbb{C}^{C \times C}$ is the CM of the input BF vector, defined as

$$\mathbf{R} = E\{\psi[n]\psi^H[n]\} \approx \frac{1}{N} \sum_{n=1}^{N} \psi[n]\psi^H[n]. \tag{7}$$

In above, $\psi[n]$ is the instantaneous input vector at time instant n (i.e., a row of $\boldsymbol{\Omega}$), and N is the total number of available samples over which $\mathbf{R}$ is estimated. If the input signal does not vary significantly within DPD iterations, the CM can be estimated off-line and then kept fixed, avoiding repetitive online calculations. For later purposes, we also already define the ordinary autocorrelation function of the DPD input signal $x[n]$, as well as its sample estimate, as

$$R_X(\tau) = E\{x[n]x^*[n-\tau]\} \approx \frac{1}{N} \sum_{n=1}^{N} x[n]x^*[n-\tau]. \tag{8}$$

In general, it is important to note that the term $(\boldsymbol{\Omega}_k^H \boldsymbol{\Omega}_k)^{-1}$ in the GN learning rule essentially calculates the (conjugate of the) ICM at every DPD iteration. This provides an accurate description of the statistics of the input signal at every iteration but also involves heavy computational complexity. For this reason, the GN method is considered as the baseline high-performance solution against which the linearization performance is compared. At the same time, the SO approach aims at drastically reducing the computational complexity by obtaining the ICM with different alternatives, described in Sections III and IV, while still achieving a linearization performance close to GN. We further note that the learning solutions presented in (5) and (6) are generic learning approaches for linear-in-parameters models and can also be applied to other models, such as GMP [3], Volterra DDR [35], [36], or even LUT-based solutions [29], [37].

Finally, it is noted that one common alternative in the DPD developments is to deploy BF orthogonalization or whitening [21], [23]. Such an approach can also facilitate utilizing low-complexity learning algorithms, even block-LMS [22]. However, the BF whitening also implies increased DPD main path complexity—even more so in case of dynamic input signals, implying that also the whitening matrix should be recalculated. Hence, in this work, we do not consider BF whitening in the DPD main path.

C. Observation Receiver Configuration in mmW Active Array Systems

In the context of mmW phased-array systems, there exist various alternatives for arranging the ORX in order to obtain the combined learning signal—$y[n]$ in Fig. 1—used for DPD parameter learning. One particular approach is the hardware-based method, in which each PA output signal is coupled, phase-aligned, and combined in hardware [11], [12], [38]. Another plausible solution is to configure a separate ORX, which captures the combined OTA received signal and feeds it back to the DPD systems for DPD learning [13], [39]–[41]. These two alternatives are presented, as plausible solutions, in Fig. 1. Both of these alternatives mimic the far-field combined signal at a distant receiver, in case of line-of-sight propagation.

It is important to clarify that the DPD learning solutions and methods presented within this article do not depend on the actual way of obtaining the combined learning signal used for parameter estimation. It is noted, however, that the hardware-based approach does not suffer from the common OTA-related challenges, such as ORX misalignment and positioning challenges, environmental dependencies, and beam dependence of radiated nonlinear distortion. We also note that there exist several works in the literature [39]–[41], where the OTA ORX beam misalignment challenges are studied, proposing different solutions to provide the far-field signal in the main beam direction. Furthermore, our adopted mmW measurement setup, featured in Section VI, aims at mimicking the hardware-based feedback system by carefully aligning a horn antenna in the main beam, acting as the OTA ORX.

III. LOW-COMPLEXITY METHODS FOR MODIFYING THE ICM

If the input signal to the SO algorithm remains static, the ICM can be precalculated off-line, saving complexity in the DPD learning path. However, whenever the input signal changes, the ICM needs to be estimated again in order to provide optimal linearization performance. The process of calculating the ICM—or, alternatively, the term $(\boldsymbol{\Omega}_k^H \boldsymbol{\Omega}_k)^{-1}$ in the GN learning rule (6)—is computationally heavy, as it involves calculating the inverse of the product between the BF matrix and its Hermitian transpose. The size of the BF matrix, $\boldsymbol{\Omega}_k$, is $K \times C$, which is typically large in DPD implementations. To ease this calculation, we propose in Sections III-A and III-B two methods for modifying the ICM when the input signal or the parameterization of the DPD is changed. We finally note that these solutions are applicable to any DPD models that are linear in the parameters, not just the MP.

A. Reducing the Dimensionality of the ICM

The first considered method avoids the need of recalculating the ICM if a simpler parameterization is desired in the DPD system. A simpler parameterization might be adopted when lowering the transmit power or when changing from a high-order modulation to a lower order modulation with less stringent EVM requirements. Furthermore, in mmW active array linearization, where the severity of the nonlinear

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

PASCUAL CAMPO *et al.*: INVERSE CM ESTIMATION FOR LOW-COMPLEXITY CLOSED-LOOP DPD SYSTEMS

5

distortions varies as a function of the beam direction, such a technique may be particularly useful in optimizing the power consumption of the transmitter.

To this end, we first generate a generic ICM obtained with a large DPD parameterization. If a smaller polynomial order or memory depth is desired, the proposed method then removes the appropriate rows and columns from the generic ICM and obtains a new ICM submatrix, which corresponds to the new chosen parameterization. The new submatrix can then be used in the SO learning equation to estimate the new DPD coefficients.

The proposed method can be described as follows. Let us consider a generic ICM, $\tilde{\mathbf{Q}}^{-1} = (\tilde{\mathbf{R}}^*)^{-1} \in \mathbb{C}^{\tilde{C} \times \tilde{C}} = (\tilde{\boldsymbol{\Omega}}^H \tilde{\boldsymbol{\Omega}})^{-1}$, obtained from the complete set of BFs, $\tilde{\boldsymbol{\Omega}}$. This matrix is known and it is used to find the reduced ICM subset $\mathbf{Q}^{-1} = (\mathbf{R}^*)^{-1} \in \mathbb{C}^{C \times C} = (\boldsymbol{\Omega}^H \boldsymbol{\Omega})^{-1}$, as described before. The BF matrices, $\tilde{\boldsymbol{\Omega}}$ and $\boldsymbol{\Omega}$, are mutually related as

$$\tilde{\boldsymbol{\Omega}} = (\boldsymbol{\Omega} \quad \mathbf{v}) \tag{10}$$

where $\mathbf{v} \in \mathbb{C}^{N \times 1}$ represents a column in $\tilde{\boldsymbol{\Omega}}$ that needs to be removed. The generic CM can be now written as

$$\tilde{\mathbf{Q}} = \tilde{\boldsymbol{\Omega}}^H \tilde{\boldsymbol{\Omega}} = \begin{pmatrix} \boldsymbol{\Omega}^H \\ \mathbf{v}^H \end{pmatrix} (\boldsymbol{\Omega} \quad \mathbf{v}) = \begin{pmatrix} \boldsymbol{\Omega}^H \boldsymbol{\Omega} & \boldsymbol{\Omega}^H \mathbf{v} \\ \mathbf{v}^H \boldsymbol{\Omega} & \mathbf{v}^H \mathbf{v} \end{pmatrix} \tag{11}$$

and its block inverse can be directly obtained by using the Schur complement [42] of (11) as

$$\tilde{\mathbf{Q}}^{-1} = \left(\begin{array}{c|c} \mathbf{A}^{-1} & \mathbf{B}^{-1} \\ \hline \mathbf{C}^{-1} & \mathbf{D}^{-1} \end{array} \right) \tag{12}$$

where

$$\mathbf{A}^{-1} = (\boldsymbol{\Omega}^H \boldsymbol{\Omega})^{-1} + \mathbf{D}^{-1} (\boldsymbol{\Omega}^H \boldsymbol{\Omega})^{-1} \boldsymbol{\Omega}^H \mathbf{v} \mathbf{v}^H \boldsymbol{\Omega} \left((\boldsymbol{\Omega}^H \boldsymbol{\Omega})^{-1} \right)^H \tag{13}$$

$$\mathbf{B}^{-1} = -\mathbf{D}^{-1} (\boldsymbol{\Omega}^H \boldsymbol{\Omega})^{-1} \boldsymbol{\Omega}^H \mathbf{v}, \tag{14}$$

$$\mathbf{C}^{-1} = -\mathbf{D}^{-1} \mathbf{v}^H \boldsymbol{\Omega} \left((\boldsymbol{\Omega}^H \boldsymbol{\Omega})^{-1} \right)^H \tag{15}$$

$$\mathbf{D}^{-1} = \left(\mathbf{v}^H \mathbf{v} - \mathbf{v}^H \boldsymbol{\Omega} (\boldsymbol{\Omega}^H \boldsymbol{\Omega})^{-1} \boldsymbol{\Omega}^H \mathbf{v} \right)^{-1}. \tag{16}$$

Finally, the system of linear equations in (13) can be directly solved to obtain $(\boldsymbol{\Omega}^H \boldsymbol{\Omega})^{-1}$, which corresponds to the desired ICM subset, expressed as

$$\mathbf{Q}^{-1} = (\mathbf{R}^*)^{-1} = (\boldsymbol{\Omega}^H \boldsymbol{\Omega})^{-1} = \mathbf{A}^{-1} - \mathbf{B}^{-1} \left(\frac{\mathbf{B}^{-1}}{\mathbf{D}^{-1}} \right)^H. \tag{17}$$

We note that removing one column from the BF matrix, $\tilde{\boldsymbol{\Omega}}$, corresponds to removing one row and one column from the ICM, $\tilde{\mathbf{Q}}^{-1}$. The proposed method requires moving the row and column to be removed to the end of the matrix and can be executed iteratively to cut out more than one row/column. In Algorithm 1, we present an example pseudocode that shows how to remove l rows and columns from $\tilde{\mathbf{Q}}^{-1} \in \mathbb{C}^{\tilde{C} \times \tilde{C}}$, obtaining the new ICM submatrix $\mathbf{Q}^{-1} \in \mathbb{C}^{(\tilde{C}-l) \times (\tilde{C}-l)}$.

Algorithm 1 Pseudocode showing how l rows and columns can be removed from $\tilde{\mathbf{Q}}^{-1}$ to obtain the subset $\mathbf{Q}^{-1}$. The generic ICM, given as input data, is conceptually presented in (12)

Data: $\tilde{\boldsymbol{\Omega}}$, $\tilde{\mathbf{Q}}^{-1}$
Result: $\mathbf{Q}^{-1}$
Initialize: $\mathbf{Q}^{-1} \in \mathbb{R}^{C \times C} = \mathbf{I}$,
         $\mathbf{A}^{-1} \in \mathbb{R}^{(\tilde{C}-1) \times (\tilde{C}-1)} = \mathbf{0}$,
         $\mathbf{B}^{-1} \in \mathbb{R}^{(\tilde{C}-1) \times 1} = \mathbf{0}$,
         $\mathbf{D}^{-1} = 0$;
for $i = 1, 2, \ldots, l-1, l$, **do**
  | Permute row/column to the end of $\tilde{\mathbf{Q}}^{-1}$;
  | $\mathbf{A}^{-1} = \tilde{\mathbf{Q}}^{-1}(1:(\tilde{C}-i), \ 1:(\tilde{C}-i))$;
  | $\mathbf{D}^{-1} = \tilde{\mathbf{Q}}^{-1}(\tilde{C}-i+1, \tilde{C}-i+1)$;
  | $\mathbf{B}^{-1} = \tilde{\mathbf{Q}}^{-1}(1:(\tilde{C}-i), \tilde{C}-i+1)$;
  | $\mathbf{Q}^{-1} = \mathbf{A}^{-1} - \mathbf{B}^{-1} (\frac{\mathbf{B}^{-1}}{\mathbf{D}^{-1}})^H$;
end

B. Frequency Shifting the ICM

In the second considered scenario, we study how to avoid the ICM recalculation when the DPD input signal is digitally frequency shifted. In such a case, the spectral components of the signal are modified, and thus, its CM and ICM will no longer be the same. In order to provide accurate linearization performance, the new ICM needs to be estimated again by some means. To that end, we propose a novel method that shifts the original ICM, denoted as $\tilde{\mathbf{Q}}^{-1} = (\tilde{\mathbf{R}}^*)^{-1}$, in the frequency domain to obtain the resulting shifted ICM, denoted as $\mathbf{Q}^{-1} = (\mathbf{R}^*)^{-1}$. The computational complexity of this method is very low compared to recalculating the whole ICM again, as shown in Section V. A good example use case of such frequency shifting is when the uplink spectral allocation of a terminal/UE transmitter changes within a given channel bandwidth.

The proposed method is constructed as follows. Let us denote the original unshifted digital input signal as $\tilde{x}[n]$ and the corresponding shifted signal as $x[n]$. A frequency shift of Δf, in Hz, can be mathematically expressed in the time domain as the multiplication with a sampled complex exponential, and thus, the shifted digital signal reads

$$x[n] = e^{j\Delta\omega n} \tilde{x}[n] \tag{19}$$

where $\Delta\omega = (2\pi \Delta f)/f_s$ denotes the normalized angular frequency shift, while f_s refers to the sample rate. The next step is to obtain the shifted BF matrix, $\boldsymbol{\Omega}$, exclusively as a function of the unshifted signal $\tilde{x}[n]$ and the normalized frequency shift $\Delta\omega$. This can be done by substituting (19) into (3) and (4). For simplicity, we just present the first row of the resulting $\boldsymbol{\Omega}$, shown in (9) at the bottom of the next page, considering that $|e^{j\Delta\omega n}| = 1$. Due to the formulation taken in (9), $\mathbf{Q}$ can also be expressed as a function of $\tilde{\mathbf{Q}}$ and a multiplying factor that depends on the frequency shift $\Delta\omega$. This factor appears in the BF terms with different delays, caused by the distinct memory terms appearing in the BF matrix. Specifically, we observe 0-sample-delay terms,

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

6

IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES

1-sample-delay terms, and, in general, m-sample-delay terms, appearing in the CM, expressed as

$$0\text{-sample-delay}: \mathbf{Q}(i, j) = e^{-j\Delta\omega n}e^{j\Delta\omega n}\tilde{\mathbf{Q}}(i, j) = \tilde{\mathbf{Q}}(i, j)$$
$$1\text{-sample-delay}: \mathbf{Q}(i, j) = e^{-j\Delta\omega n}e^{j\Delta\omega(n-1)}\tilde{\mathbf{Q}}(i, j)$$
$$= e^{-j\Delta\omega 1}\tilde{\mathbf{Q}}(i, j)$$
$$\vdots$$
$$m\text{-sample-delay}: \mathbf{Q}(i, j) = \cdots = e^{-j\Delta\omega m}\tilde{\mathbf{Q}}(i, j). \quad (20)$$

The CM is generally a Hermitian–Toeplitz matrix, and the way how the different sample delay terms are scattered through it can also be seen from (28), where the value of τ within the autocorrelation function $R_X(\tau)$, defined formally in (8), indicates the corresponding delay. Hence, the zero-delay samples correspond to the main diagonal of the $M \times M$ submatrices within the CM, the one-delay samples correspond to the second diagonal term of these submatrices, and so on. From this structure, it is then relatively easy to see that the frequency-shifted CM can be written as

$$\mathbf{Q} = \mathbf{P}^H\tilde{\mathbf{Q}}\mathbf{P} \quad (21)$$

where the diagonal matrix $\mathbf{P} \in \mathbb{C}^{C \times C}$ applies the right shift weights for each column of $\tilde{\mathbf{Q}}$, and it is defined as

$$\mathbf{P} = \begin{bmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & e^{j\Delta\omega m} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & e^{j\Delta\omega m} \end{bmatrix}. \quad (22)$$

Defining the CM as in (21) will allow for applying simple yet efficient matrix inverse properties. As a result, the shifted ICM can be directly obtained from (21) as

$$\mathbf{Q}^{-1} = \left(\mathbf{R}^*\right)^{-1} = \mathbf{P}^{-1}\tilde{\mathbf{Q}}^{-1}\mathbf{P}^{H(-1)}. \quad (23)$$

Note that only the original ICM, $\tilde{\mathbf{R}}^{-1}$, and the frequency shift, $\Delta\omega$, are needed to obtain the final shifted ICM, $\mathbf{R}^{-1}$.

IV. Low-Complexity ICM Estimation for Gaussian Signals

This section presents two novel methods to estimate the ICM in a computationally efficient manner. These methods build on the autocorrelation function of the input signal and provide a set of precomputed coefficients that can be utilized to easily estimate the ICM. The first method is an application of Reed's theorem [43] for calculating the CM, while the second technique utilizes the Bussgang theorem and other stochastic analysis to provide an approximate CM. In the latter case, the resulting CM can be shown to consist of a specific subblock structure expressible as a Kronecker product, allowing for efficient inversion. We note that, since we consider complex Gaussian process theory, we assume the Gaussian distributed input signals in this section. OFDM signals are known to converge toward Gaussian, as the number of subcarriers is large [44]. This claim is further substantiated in Section VI.

A. ICM Estimation With the Autocorrelation Function

The first proposed method utilizes the moment theorem for complex-circular Gaussian signals presented in [43]. We specifically aim at expressing the CM as a function of the autocorrelation function of the input signal. To that end, the autocorrelation function is first estimated over the desired span with (8). Reed's theorem states that all odd-order central product moments are zero, while all even-order moments can

$$\boldsymbol{\Omega} = \left(x[n]e^{j\Delta\omega n} \cdots x[n]|x[n]|^{(P-1)}e^{j\Delta\omega n} \quad x[n-1]e^{j\Delta\omega(n-1)} \cdots x[n-1]|x[n-1]|^{(P-1)}e^{j\Delta\omega(n-1)} \cdots \right.$$
$$\left. \cdots x[n-M+1]e^{j\Delta\omega(n-M+1)} \cdots x[n-M+1]|x[n-M+1]|^{(P-1)}e^{j\Delta\omega(n-M+1)}\right) \quad (9)$$

$$E\left\{x[n]x^*[n-1]|x[n-1]|^2\right\} = E\left\{x^*[n-1]x[n]\right\}E\left\{x^*[n-1]x[n-1]\right\} + E\left\{x^*[n-1]x[n]\right\}E\left\{x^*[n-1]x[n-1]\right\}$$
$$= 2R_X(\tau=0)R_X(\tau=1),$$

$$E\left\{x[n-1]x^*[n]|x[n]|^2\right\} = E\left\{x^*[n]x[n]\right\}E\left\{x^*[n]x[n-1]\right\} + E\left\{x^*[n]x[n]\right\}E\left\{x^*[n]x[n-1]\right\}$$
$$= 2R_X(\tau=0)R_X^*(\tau=1),$$

$$E\left\{x[n]x^*[n-1]|x[n]|^2|x[n-1]|^2\right\} = E\left\{x^*[n]x[n]\right\}E\left\{x^*[n-1]x[n]\right\}E\left\{x^*[n-1]x[n-1]\right\}$$
$$+ E\left\{x^*[n]x[n]\right\}E\left\{x^*[n-1]x[n]\right\}E\left\{x^*[n-1]x[n-1]\right\}$$
$$+ E\left\{x^*[n-1]x[n]\right\}E\left\{x^*[n]x[n]\right\}E\left\{x^*[n-1]x[n-1]\right\}$$
$$+ E\left\{x^*[n-1]x[n]\right\}E\left\{x^*[n-1]x[n]\right\}E\left\{x^*[n]x[n-1]\right\}$$
$$+ E\left\{x^*[n-1]x[n]\right\}E\left\{x^*[n]x[n]\right\}E\left\{x^*[n-1]x[n-1]\right\}$$
$$+ E\left\{x^*[n-1]x[n]\right\}E\left\{x^*[n-1]x[n]\right\}E\left\{x^*[n]x[n-1]\right\}$$
$$= 4R_X^2(\tau=0)R_X(\tau=1) + 2R_X^2(\tau=1)R_X^*(\tau=1),$$

$$E\left\{x[n-1]x^*[n]|x[n]|^2|x[n-1]|^2\right\} = \cdots = 4R_X^2(\tau=0)R_X^*(\tau=1) + 2\left(R_X^*(\tau=1)\right)^2 R_X(\tau=1) \quad (18)$$

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

PASCUAL CAMPO *et al.*: INVERSE CM ESTIMATION FOR LOW-COMPLEXITY CLOSED-LOOP DPD SYSTEMS

7

be obtained through sums of products of the second-order moments, expressed in the general case as [43]

$$E\{x_0^* x_1^* \cdots x_{k-1}^* x_0 x_1 \cdots x_{k-1}\}$$
$$= \sum_\zeta E\{x_\zeta^* x_0\} E\{x_\zeta^* x_1\} \cdots E\{x_\zeta^* x_{k-1}\}. \quad (25)$$

Here, for notational simplicity, $x_m = x[n - m]$, and ζ constitutes a permutation of the set $\{0, 1, \ldots, (k-1)\}$. Noting that all terms in the CM have an even order, and following the specific BF order assumed in (3), each term in the CM can be obtained by applying (25).

For clarity, an example of the application of Reed's theorem to the fourth- and sixth-order moments is presented in (18), as shown at the bottom of the previous page, while other higher order terms can be obtained similarly. The definition of the autocorrelation function in (8) has been used to arrive at the last expressions in (18). It is noted that the expression presented in (25) allows to obtain the higher order terms recursively. It is also noted that the specific terms of the form $E\{|x[n]|^{2p}\}$ can be alternatively obtained as

$$E\{|x[n]|^{2p}\} = p! E\{|x[n]|^2\}^p = p! \sigma_x^{2p} \quad (26)$$

which may simplify their calculation.

Once Reed's theorem has been applied to each individual term in the CM, we can express the CM with a specific sub-block matrix structure. Specifically, the CM can be expressed with a $\lceil (P/2) \rceil \times \lceil (P/2) \rceil$ submatrix structure, each of size $M \times M$, written as presented in (24), as shown at the bottom of the page. Here, the submatrix $\mathbf{R}_{k,l}$ is in turn defined as

$$\mathbf{R}_{k,l} = E\left\{\begin{bmatrix} \psi_k[n]\psi_l^*[n] & \cdots & \psi_k[n]\psi_{l+M-1}^*[n] \\ \vdots & \ddots & \vdots \\ \psi_{k+M-1}[n]\psi_l^*[n] & \cdots & \psi_{k+M-1}[n]\psi_{l+M-1}^*[n] \end{bmatrix}\right\} \quad (27)$$

where the subscript in $\psi[n]$ indicates the corresponding element within the instantaneous input vector. Due to the notation utilized in (24) and (27), such matrices can accommodate any polynomial order, P, and memory depth, M, chosen in the DPD system. Note that only the calculation of the upper triangular submatrices in $\mathbf{R}$ is required to build the complete CM. The CM is then inverted to generate the resulting ICM, thereon used in the SO learning equation.

B. ICM Approximation With Bussgang Coefficients

In the second proposed method, we aim at simplifying the previous method by using Bussgang's decomposition. This theorem states that the cross correlation of a Gaussian signal $x[n]$ and a signal $y[n] = f(x[n])$ that has passed through an instantaneous nonlinear function $f(\cdot)$ can be expressed as the product of the autocorrelation function of $x[n]$ and a scaling constant. Formally, this is expressed as

$$R_{XY}(\tau) = \xi R_X(\tau) \quad (29)$$

where

$$R_{XY}(\tau) = E\{x[n]y^*[n - \tau]\} \quad (30)$$

is the definition of the cross correlation function, while the autocorrelation function $R_X(\tau)$ reads as in (8). The Bussgang coefficient, ξ, is obtained through the complex proper Gaussian probability distribution function (PDF) as

$$\xi = \frac{1}{\pi \sigma_x^4} \int_{-\infty}^{\infty} x^* f(x) e^{-\frac{|x|^2}{\sigma_x^2}} dx \quad (31)$$

where $f(\cdot)$ represents the nonlinear amplitude distortion and σ_x^2 is the variance of $x[n]$.

In general, the second-order terms appearing in the CM, on its first row, can be expressed directly as a function of $R_X(\tau)$ as

$$\mathbf{R}(1, 1) = R_X(\tau = 0) \quad (32)$$
$$\mathbf{R}(1, 2) = R_X(\tau = 1) \quad (33)$$
$$\vdots$$
$$\mathbf{R}(1, M) = R_X(\tau = M). \quad (34)$$

The higher order terms appearing in the CM (i.e., $x[n]|x[n]|^p$, $p = 2, 4, \ldots$) can be seen as nonlinear functions of $x[n]$ and can thus be expressed as a function of the autocorrelation according to the Bussgang theorem [45].

Using (29), we can now express all the remaining high-order terms in the CM uniquely as a function of $R_X(\tau)$ as

$$E\{x[n]x^*[n]|x[n]|^2\} = \xi_1 R_X(\tau = 0) \quad (35)$$
$$E\{x[n]x^*[n - 1]|x[n - 1]|^2\} = \xi_1 R_X(\tau = 1) \quad (36)$$
$$E\{x[n]x^*[n]|x[n]|^4\} = \xi_2 R_X(\tau = 0) \quad (37)$$
$$E\{x[n]x^*[n - 1]|x[n - 1]|^4\} = \xi_2 R_X(\tau = 1) \quad (38)$$
$$\vdots$$

where the coefficients $\xi_1, \xi_2, \ldots$ can be obtained using (31). By having all the terms as a function of $R_X(\tau)$ and the Bussgang coefficients, the complete CM can be expressed through them. A concrete example is presented in (28), as shown at the bottom of the next page, which is showing the CM structure in the specific case of $P = 3$ and $M = 1$ while also assuming that $\xi_0 = 1$. Each matrix element, which depends on the input data signal, can be then obtained by using the explicit expressions shown in (7) and (31).

We next note that the order of the BFs in the BF matrix becomes important to replicate this particular structure. By further studying (28), we realize that it can be formulated with a

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{1,1} & \mathbf{R}_{1,M+1} & \cdots & \mathbf{R}_{1,(\lceil \frac{P}{2} \rceil - 1)M+1} \\ \mathbf{R}_{1,M+1} & \mathbf{R}_{M+1,M+1} & \cdots & \mathbf{R}_{M+1,(\lceil \frac{P}{2} \rceil - 1)M+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{1,(\lceil \frac{P}{2} \rceil - 1)M+1} & \mathbf{R}_{M+1,(\lceil \frac{P}{2} \rceil - 1)M+1} & \cdots & \mathbf{R}_{(\lceil \frac{P}{2} \rceil - 1)M+1,(\lceil \frac{P}{2} \rceil - 1)M+1} \end{bmatrix} \quad (24)$$

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

8   IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES

specific subblock matrix notation that will allow for efficient Kronecker inversion. The CM can be alternatively simplified in $M \times M$ subblocks, which are weighted with a different Bussgang coefficient and are propagated through $\mathbf{R}$ as

$$
\mathbf{R} = \begin{bmatrix} \mathbf{R}_0 & \mathbf{R}_1 & \mathbf{R}_2 & \cdots & \mathbf{R}_{\lfloor \frac{P}{2} \rfloor} \\ \mathbf{R}_1 & \mathbf{R}_2 & \mathbf{R}_3 & \cdots & \mathbf{R}_{\lfloor \frac{P}{2} \rfloor + 1} \\ \mathbf{R}_2 & \mathbf{R}_3 & \mathbf{R}_4 & \cdots & \mathbf{R}_{\lfloor \frac{P}{2} \rfloor + 2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{\lfloor \frac{P}{2} \rfloor} & \mathbf{R}_{\lfloor \frac{P}{2} \rfloor + 1} & \mathbf{R}_{\lfloor \frac{P}{2} \rfloor + 2} & \cdots & \mathbf{R}_{2\lfloor \frac{P}{2} \rfloor} \end{bmatrix}
$$

(39)

where the subscripts indicate the corresponding Bussgang coefficients. In addition, the submatrix $\mathbf{R}_k \in \mathbb{C}^{M \times M}$ reads

$$
\mathbf{R}_k = \xi_k \begin{bmatrix} R_X(\tau = 0) & R_X(\tau = 1) & \cdots & R_X(\tau = M) \\ R_X^*(\tau = 1) & R_X(\tau = 0) & \cdots & R_X(\tau = M - 1) \\ \vdots & \vdots & \ddots & \vdots \\ R_X^*(\tau = M) & R_X^*(\tau = M - 1) & \cdots & R_X(\tau = 0) \end{bmatrix}.
$$

(40)

Note that only the submatrices appearing on the first row and the last column of (39) need to be calculated to build the whole CM.

Finally, due to the well-structured formulation in (39), the CM can be efficiently inverted using different matrix inversion methods. One possible solution is to use block-recursive matrix inverse algorithms, such as the one in [46], which poses drastically reduced complexity when compared to ordinary matrix inversion. An alternative and yet more efficient solution is to use the Kronecker inversion [47], [48]. To this end, we first note that the CM can be equivalently expressed with the Kronecker product as

$$
\mathbf{R} = \Xi \otimes \mathbf{R}_0 \tag{41}
$$

where $\Xi \in \mathbb{R}^{\lceil (P/2) \rceil \times \lceil (P/2) \rceil}$ contains the complete set of Bussgang coefficients. Then, the ICM is directly obtained as the inverse of its elements as

$$
\mathbf{R}^{-1} = \Xi^{-1} \otimes \mathbf{R}_0^{-1}. \tag{42}
$$

The Kronecker inversion further reduces the computational complexity of inverting $\mathbf{R}$, as demonstrated in Section V.

V. Complexity Analysis and Comparison

In this section, we present the computational complexity involved in the proposed DPD and ICM estimation solutions. The processing complexity is divided in three different stages—the DPD main path (Section II-A), the DPD learning path (Section II-B), and the ICM-related calculations (Sections III and IV). All the analyses are carried out in terms of real multiplications per K-sized DPD iteration (i.e.,

for one block iteration with data size of K samples). Multiplications constitute, in general, the most resource-intensive operations in digital signal processing (DSP) implementations, while additions can be considered to be essentially free [15], [49]. In this study, it is assumed that one complex-complex multiplication is calculated with four real multiplications, and one real-complex multiplication is calculated with two real multiplications. In addition, the matrix inversion of an $m \times m$ matrix is assumed to cost 4 m^3 real multiplications [42], [46].

As noted, the first considered stage is the DPD main path, where the actual predistortion is applied. Here, the complete BF matrix [i.e., Ω in (3)] is first calculated from the input signal $x[n]$, assuming its recursive calculation through previously obtained values. Later, the actual predistortion processing is carried out as in (2). The second considered stage is the DPD learning path, where the closed-loop learning equations are applied to update the DPD coefficients. Here, we consider both the reference GN algorithm and the SO learning approach. In this analysis, in case of SO learning, we express the learning complexity first for a given ICM while noting the ICM calculation complexity then separately. The GN approach, in turn, needs to calculate the ICM estimate $(\Omega_k^H \Omega_k)^{-1}$ in every DPD iteration—by its very definition. Finally, a complexity assessment and comparison between the classical ICM calculation and the proposed ICM estimation methods are provided. The classical ICM estimation complexity directly results from the calculation of $(\Omega^H \Omega)^{-1}$. The complexity of the proposed ICM estimation methods, in turn, is determined following the exact processing principles presented in Sections III and IV.

Table I gathers the obtained complexity analysis results for the considered methods. The table presents the required real multiplications in each stage of the DPD solution, in a general form as a function of the modeling parameters. In addition, the final column presents a specific numerical example when $P = 9$, $M = 4$, $C = 20$, $l = 5$, and $K = 20\,000$, in order to provide a concrete and representative example of the involved complexities. Such DPD parameterization is similar to the one chosen to perform the RF experiments and validation in Section VI. From the numerical example, it can be clearly observed how the computational complexity is reduced against the reference GN method when utilizing the SO learning solution in combination with the proposed ICM estimation methods. In all cases, the learning complexity is drastically reduced while still achieving a similar linearization performance, as demonstrated in Section VI. The quantitative results also demonstrate the very large reduction in the needed multiplications, achievable through the proposed ICM estimation methods, compared to the direct ICM calculation.

$$
\mathbf{R} = \begin{bmatrix} \xi_0 R_X(\tau = 0) & \xi_0 R_X(\tau = 1) & \xi_1 R_X(\tau = 0) & \xi_1 R_X(\tau = 1) \\ \xi_0 R_X^*(\tau = 1) & \xi_0 R_X(\tau = 0) & \xi_1 R_X^*(\tau = 1) & \xi_1 R_X(\tau = 0) \\ \xi_1 R_X(\tau = 0) & \xi_1 R_X(\tau = 1) & \xi_2 R_X(\tau = 0) & \xi_2 R_X(\tau = 1) \\ \xi_1 R_X^*(\tau = 1) & \xi_1 R_X(\tau = 0) & \xi_2 R_X^*(\tau = 1) & \xi_2 R_X(\tau = 0) \end{bmatrix}
$$

(28)

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

PASCUAL CAMPO *et al.*: INVERSE CM ESTIMATION FOR LOW-COMPLEXITY CLOSED-LOOP DPD SYSTEMS 9

TABLE I

COMPUTATIONAL COMPLEXITIES OF THE DPD MAIN PATH, DPD LEARNING PATH, AND THE VARIOUS ICM ESTIMATION METHODS

Operation		Amount of real multiplications	Numerical example with $P = 9$, $M = 4$, $C = 20$ $l = 5$, $K = 20,000$
DPD main path	BF generation	$2K(\lfloor \frac{P}{2} \rfloor + 1)$	200,000
	Digital predistortion	$4KC$	1,600,000
	Total	$2K(\lfloor \frac{P}{2} \rfloor + 2C + 1)$	1,800,000
DPD learning path	Damped Gauss-Newton	$4C(C^2 + CK + C + K + \frac{1}{2})$	33,633,640
	Self-orthogonalization	$4C(K + C + \frac{1}{2})$	1,601,640
Other: ICM calculation (only when required)	Classical ICM calculation	$4C^2(C + K)$	32,032,000
	Dimensionality red. ICM (Sec. III-A)	$4lC(C - 1)$	7,600
	Frequency shifting ICM (Sec. III-B)	$(C^2 - C)/2$	190
	Autocorr. ICM est. (Sec. IV-A)	$(\lfloor \frac{P}{2} \rfloor^2 + 3\lfloor \frac{P}{2} \rfloor + 2)3M^2 + 4(C^3(M + 1)(K + 1))$	433,460
	Bussgang's ICM est. (Sec. IV-B)	$M(4M^2 + 6M\lceil \frac{P}{2} \rceil - 3M + 4K + 4) + 4(\lceil \frac{P}{2} \rceil + K + 1)$	401,208

For reference and comparison purposes, we also address the complexity of a classical MP model with BF prewhitening or orthogonalization while considering block-LMS as the DPD coefficient update rule. With block-LMS, BF prewhitening becomes a necessary ingredient stemming from the large eigenvalue spread of the input signal CM. This process ensures that reasonably fast and stable learning convergence can be obtained by the simple LMS algorithm, which essentially uses a diagonal matrix to approximate the ICM [22]. The BF prewhitening is done in the DPD main path and is assumed here to build on the well-known Cholesky decomposition-based method [23], [42], which provides an upper triangular orthogonalization matrix. The overall main path complexity of such an approach is then the one presented on the fourth row of Table I plus the extra cost of the prewhitening stage that is a matrix–matrix product. Due to the triangular orthogonalization matrix, the extra complexity is of the form $4K(C(C + 1)/2)$. In the numerical example case of Table I, the corresponding overall main path complexity is 18.6M real multiplications, reflecting a substantial increase in comparison to the methods that do not require prewhitening. On the other hand, the cost of updating the DPD coefficients in the block-LMS learning rule [50] is given by $4CK + 2C$, which leads to $1\,600\,040$ real multiplications—a number that is of the same complexity order as the SO solution when combined with any of the proposed ICM estimation methods. This analysis thus shows that the main path complexity increases substantially when BF orthogonalization is deployed, while the corresponding decrease in the learning complexity through using block-LMS is fairly minor when the ICM calculations in the reference SO solutions build on the proposed methods. In addition, we note that in these complexity calculations, the prewhitening matrix has been considered to be precalculated off-line. However, when considering DPD applications for example in terminal transmitters, where the transmitted waveforms are very dynamic (the resource block allocation and/or output power can change on a per-slot basis), such an assumption may not be feasible. This, in turn, means that in such dynamic scenarios,

also the prewhitening matrix may need to be recalculated online during the learning procedures, further increasing the corresponding online processing burden.

Finally, based on the results in Table I, we separate three possible alternatives regarding the SO learning approach in combination with the ICM estimation.

1) The SO learning rule is applied in a dynamic waveform scenario using classical ICM calculation based on the sample estimate of the CM in (7). The ICM is thus calculated each time the waveform changes. If this happens frequently, the complexity approaches that of the GN.

2) The SO learning rule is applied in a dynamic waveform scenario using the proposed ICM estimation methods. Computational complexity may be drastically reduced compared to the previous when waveform changes are frequent.

3) The SO learning rule is applied in a static waveform scenario. In this case, the ICM is fixed and only needs to be computed once. Thus, the overall complexity is the smallest.

VI. RF MEASUREMENTS

In order to test, verify, and validate the proposed methods and algorithms, an extensive set of RF measurements is provided, in the context of an FR-2 mmW device. The deployed system features a state-of-the-art 28 GHz 64-element active phased array, acting as the nonlinear element which the proposed methods aim at linearizing. The measured 1-dB compression point of the array is $+41$ dBm, and the operating FR-2 band is 5G NR band n257. Once the measurement results are obtained, they are analyzed together with the quantitative complexity analysis results provided in Section V, to assess the performance–complexity tradeoffs of the proposed solutions.

In the context of mmW measurements with multiple-element phased arrays, several issues are to be noted. First, in an R-antenna array, there are also R parallel PAs, one per antenna unit. Each of these amplifiers will basically have
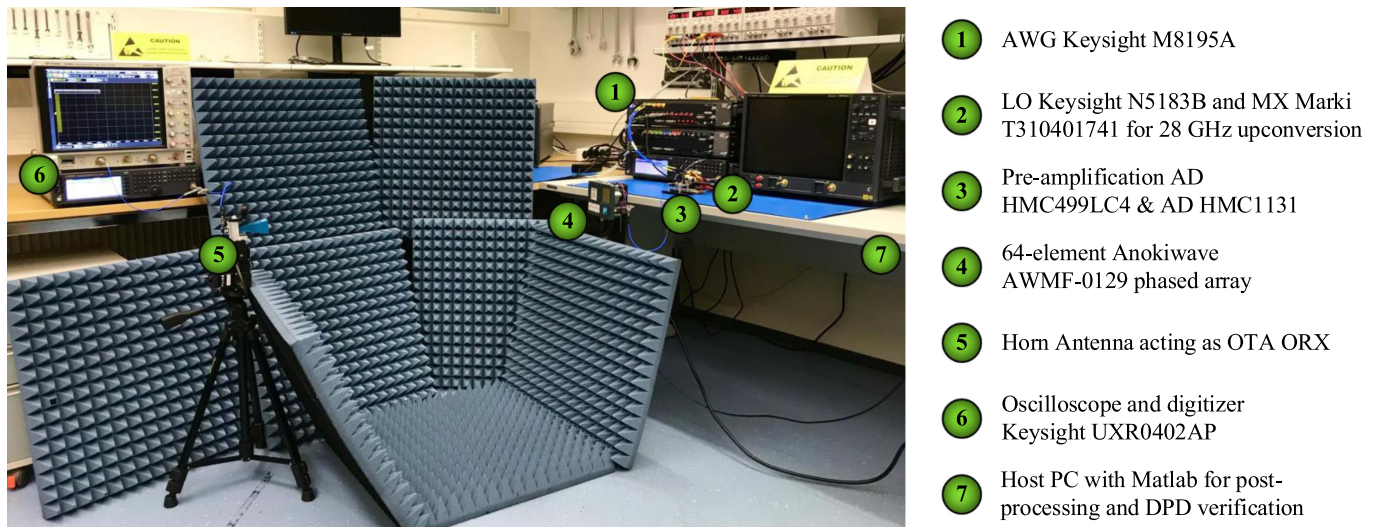
1 — AWG Keysight M8195A

2 — LO Keysight N5183B and MX Marki T310401741 for 28 GHz upconversion

3 — Pre-amplification AD HMC499LC4 & AD HMC1131

4 — 64-element Anokiwave AWMF-0129 phased array

5 — Horn Antenna acting as OTA ORX

6 — Oscilloscope and digitizer Keysight UXR0402AP

7 — Host PC with Matlab for post-processing and DPD verification

Fig. 2.   5G NR mmW/FR-2 OTA measurement setup utilized to carry out the experiments at 28 GHz.

unique nonlinear characteristics, and hence, the predistorter can only guarantee an efficient linearization performance in the main beam direction. In the rest of the angles, the intrinsic beampattern of the array suppresses the OoB nonlinear distortion [11]. Another concern is the load modulation of the PAs, occurring due to coupling between the antenna elements [8]. This makes the effective nonlinear characteristics of the antenna array beam-dependent, thus tying the DPD solution to the beam direction. Consequently, linearization methods must take considered. This issue can be tackled with, for instance, real-time DPD tracking, which is able to estimate the DPD coefficients as the beam is steered. Since the parameter estimation needs to be done in real time, it is crucial to explore ways of minimizing the involved computational complexity while at the same time being able to provide the needed linearization performance.

In this work, the DPD performance is evaluated through the well-known EVM [15], normalized mean square error (NMSE) [15], and the TRP-ACLR metrics since an OTA system is considered [7]. The TRP-ACLR measures the OoB performance by computing the ratio between the filtered mean power centered on the assigned channel frequency and the filtered mean power centered on an adjacent channel frequency, measured by integrating the powers over the whole beamspace while keeping the beamforming angle fixed [7], [8].

A. 28 GHz Active Array Experimental Setup

The complete 28 GHz measurement setup is shown in Fig. 2. The experimental setup is configured as follows. First, a Keysight M8195A arbitrary waveform generator (AWG) is deployed to output the modulated I/Q waveform at 2.5 GHz intermediate frequency (IF). Then, a Keysight N5183B signal generator provides the LO signal at 24.5 GHz that together with a Marki Microwave T310401741 mixer further upconverts the signal to the 28 GHz band. A Marki Microwave FB3300 bandpass filter (BPF) is applied immediately afterward to suppress the mixer-induced image frequencies. Two

preamplifiers, AD HMC499LC4 and AD HMC1131, are deployed to guarantee a sufficiently high power at the input of the active antenna array such that it can be pushed close toward saturation. The 64-element active array—the Anokiwave AWMF-0129—is then configured to transmit the signal OTA toward the horn antenna of the ORX. Both antennas are mounted on different electrical tripods capable of providing the required rotation in azimuth and elevation. For DPD learning and verification, both antennas are perfectly aligned when transmitting/receiving, as noted in Section II-C. For simplicity, the beamforming angle of the TX antenna array is set at 0°. The combined learning signal captured by the ORX is then attenuated and fed back to a Keysight UXR0402AP oscilloscope, which is acting as the digitizer. Finally, the digital signal is sent to a host PC for further processing and/or performance assessments. We note that the Anokiwave AWMF-0129 does not allow for actual hardware-based combiners for feedback, and hence, we adopt the carefully aligned OTA ORX to mimic such hardware-based processing.

The modulated signals adopted in the following experiments are 3GPP 5G NR Release-15 FR-2 compliant OFDM waveforms, with the subcarrier spacing (SCS) and RB allocation specified in each particular experiment. The sampling rate of the signals and DPD execution is 2 GHz. As an additional ingredient, which aims at pushing the performance boundaries of the DPD system, we also consider a wider, nonstandard-compliant, signal bandwidth of 800 MHz in some of the experiments. This is obtained by doubling the number of active subcarriers and also the OFDM waveform processing FFT size, compared to the standard-compliant 400 MHz signal. In addition, the initial PAPR of the generated signals is approximately 12 dB when measured at the 0.01% point of the instantaneous PAPR complementary cumulative distribution function (CCDF) [2]. This value is then limited through well-known iterative clipping and filtering-based processing to 8 dB, measured also at the 0.01% point of the instantaneous PAPR CCDF. Clipping the waveform, however, strictly speaking degaussianizes the input signal distribution, as it basically

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

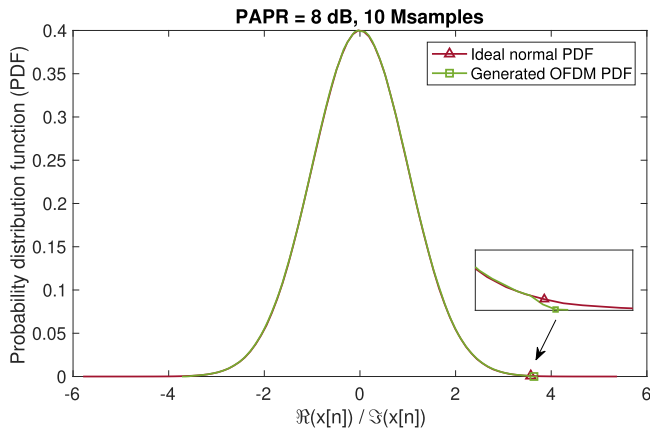PASCUAL CAMPO *et al.*: INVERSE CM ESTIMATION FOR LOW-COMPLEXITY CLOSED-LOOP DPD SYSTEMS 11

Fig. 3. PDF of the generated 5G NR FR-2 OFDM waveform with 400 MHz carrier bandwidth, including also PAPR mitigation to 8 dB through iterative clipping and filtering. The shown PDF reflects the real part of the complex I/Q waveform, while that of the imaginary part is essentially identical. For reference, also the theoretical normal PDF is shown.

removes the high- and low-end values of the amplitude PDF. The difference to the Gaussian distribution is nevertheless very small, as can be observed through the illustration in Fig. 3, which shows the PDF of the generated NR-compliant OFDM signal with clipping and filtering versus an ideal Gaussian PDF. Finally, after the PAPR mitigation, additional time-domain windowing is also applied to suppress the inherent OFDM signal sidelobes.

In each DPD iteration, a block of $K = 20\,000$ pseudorandomly generated OFDM signal samples is circularly transmitted, received, and utilized to estimate the DPD coefficients. The K-sized closed-loop error signal vector, $\mathbf{e}_k$, is generated as the subtraction between the transmitted and received data samples and essentially contains the prevailing PA distortion samples. This term is then used as an input of (5) or (6) to estimate the DPD coefficients. Such transmit/receive process is then repeated until the algorithms reach convergence. In all the experiments, the MP DPD model is configured with a ninth-order polynomial (i.e., $P = 9$) and with four taps of memory (i.e., $M = 4$). The DPD coefficients are initialized as $\boldsymbol{\alpha} = [1\,0\cdots 0]^T$ such that just the linear term is obtained after predistortion in the first iteration. In the measurements where the ICM is precalculated off-line, the CM is estimated from a 500 ksamples sequence and inverted before the DPD processing. It will remain fixed for the rest of the iterations unless otherwise mentioned.

B. Baseline DPD Performance

In this section, we study the linearization performance of the SO and the reference GN learning methods. The measurement results are carried out with both signal bandwidths of 800 MHz (120 kHz SCS, 528 RBs) and 400 MHz (120 kHz SCS, 264 RBs), measured at a highly nonlinear operation point of the active phased array. Specifically, this experiment is performed with effective isotropic radiated power (EIRP) of approximately +40.5 dBm, corresponding to an input power of approximately −6 dBm. This leads to an initial TRP-ACLR

of approximately +22 dBc and an EVM of some 11.6%, reflecting a highly nonlinear operation point.

First, Fig. 4(a) presents the obtained linearization results with a signal bandwidth of 800 MHz for both SO and reference GN learning methods. We assume for the SO method a precalculated ICM, which is then kept fixed through the DPD iterations. The GN learning rule, in contrast, calculates the ICM estimate $(\boldsymbol{\Omega}_k^H \boldsymbol{\Omega}_k)^{-1}$ in every DPD iteration. As can be seen from the figure, the linearization performance of the SO method is practically matching that of the GN, with TRP-ACLR numbers of about +32.5 dBc and EVM values below 6%. This performance is also achieved with the SO solution regardless of the rough ICM off-line estimation. Secondly, Fig. 4(b) presents the results with a signal bandwidth of 400 MHz. Similar conclusions can be drawn regarding the behavior of the SO and GN learning methods. In this case, the TRP-ACLR is around +35 dBc and the EVM is below 5.5% for both SO and GN methods. What is more important, we also include a third measurement that shows the linearization behavior of the SO learning method when the ICM is calculated from the previous 800 MHz signal (i.e., the same as the one utilized in Fig. 4(a)). As shown in the figure, the DPD performance is only slightly degraded in comparison to the normal SO learning rule, with the TRP-ACLR and EVM being around +34 dBc and 5.7%, respectively. This is because the ICM has been obtained from a wider signal than the one being transmitted. However, the SO with an ICM obtained from the 400 MHz signal would not work to linearize the wider 800 MHz signal. We finally note that all measured methods satisfy the TRP-ACLR limit of +28 dBc and EVM limit of 8% (for 64-QAM), as stated in 3GPP specifications [7].

C. Reducing the ICM Dimensionality

In this second experiment, we study and highlight the obtained numerical precision of applying the method, which removes rows/columns from the ICM to get a new covariance set for a DPD system with reduced parameterization. We recall that the rows/columns that are to be removed from the ICM are shifted to the last row/column, and then, this algorithm is repeated iteratively until all the targeted elements are successfully removed.

To this end, we perform the following experiment. We first generate an ICM of size 36×36 corresponding to $P = 11$ and $M = 6$. We then execute the proposed algorithm described in Section III-A iteratively until $l = 16$ rows and columns are removed, leading to a final ICM of size 20×20 elements—corresponding to $P = 9$ and $M = 4$. Parallel to that, we mathematically calculate, using (7), the corresponding ICM of size 20×20 (also corresponding to $P = 9$ and $M = 4$) and compare it to the estimated version. The obtained numerical error when comparing the two versions is in the order of 10^{-13}. This order of magnitude can be considered negligible in any DPD implementation, thus demonstrating the effectiveness of the considered technique. A similar experiment is next repeated for reducing the DPD parameterization down to $P = 7$ and $M = 3$, corresponding to $l = 24$. The numerical error in this case between the ideal and
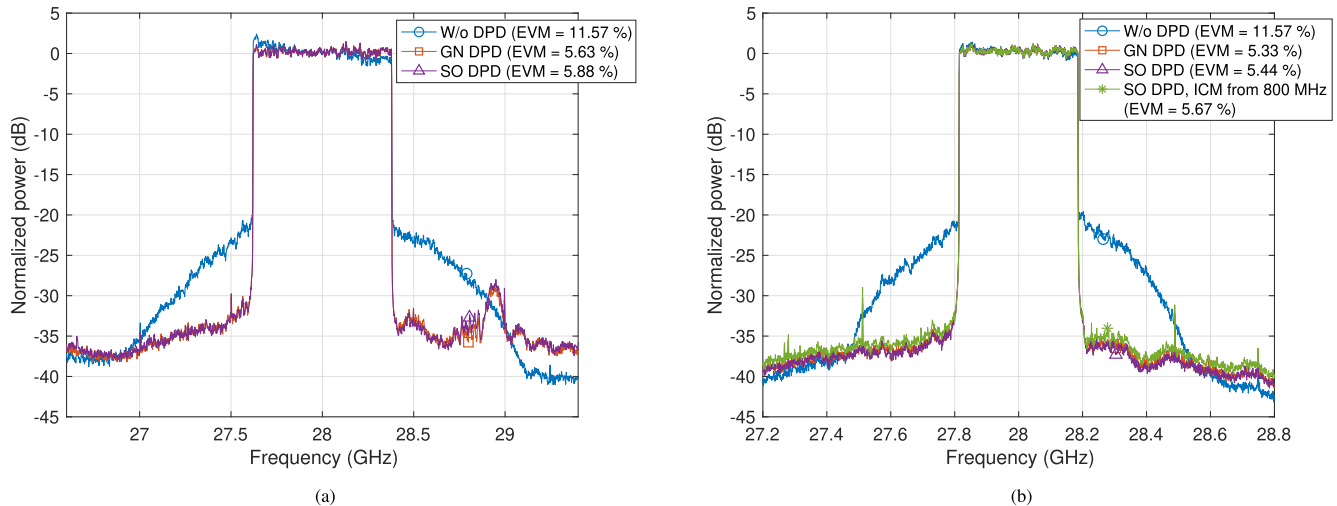
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

12   IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES

Fig. 4. Baseline OTA linearization results at EIRP $\approx$ +40.5 dBm for the SO and GN learning methods with (a) 800 MHz bandwidth and (b) 5G-NR 400 MHz signal bandwidth, also showing the SO learning rule with the ICM estimated from the 800 MHz signal.
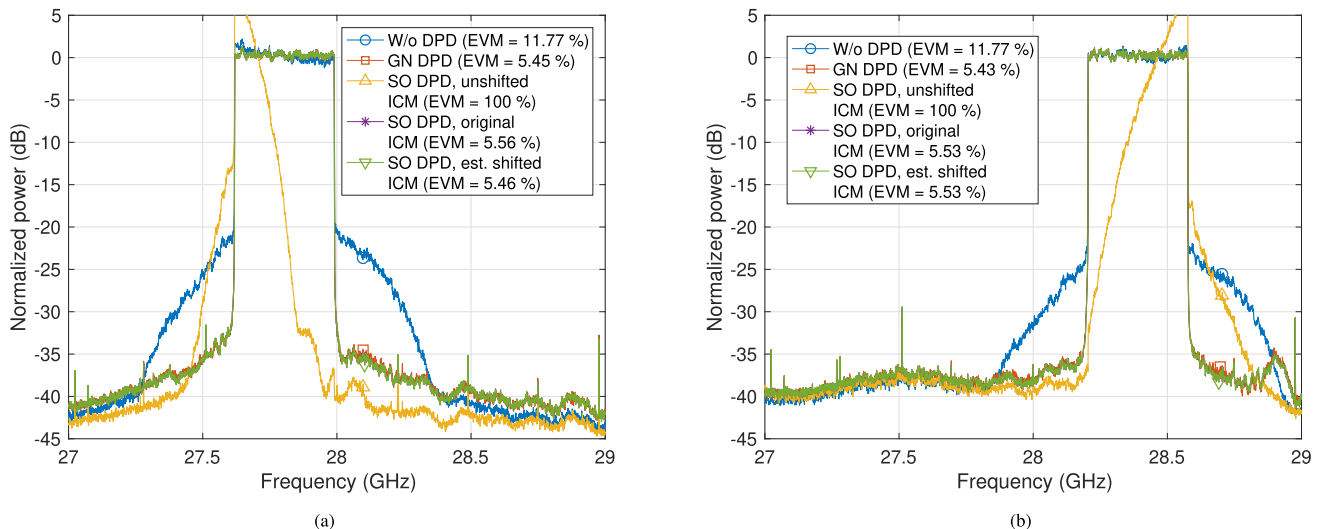

Fig. 5. 400 MHz 5G-NR OTA linearization performance at EIRP $\approx$ +40.5 dBm for the SO method, with emphasis on the ICM shifting technique proposed in Section III-B, when a frequency shift of (a) -200 and (b) $+400$ MHz is applied. The linearization performance of the GN method is also included for reference. Note that the "SO DPD, original ICM" and "SO DPD, est. shifted ICM" curves practically overlap.

estimated ICMs is in the order of 10^{-11}, reflecting again very high accuracy. Finally, we note that no PSDs are visualized along this experiment, as with such very low numerical errors, the PSDs corresponding to the calculated and estimated ICMs are essentially completely overlapping.

D. Frequency Shifting the ICM

In this third experiment, we study the linearization performance of the SO learning rule in combination with the ICM shifting method proposed in Section III-B. The measurements are carried out with a signal bandwidth of 400 MHz (120-kHz SCS, 264 RBs), while the EIRP, the array's input power, and thus initial TRP-ACLR and initial EVM values are maintained as before.

The experiments to test the proposed method are carried out as follows. First, the input 5G NR 400 MHz baseband signal is generated, and then, two different example digital

frequency shifts of -200 and $+400$ MHz are applied to it. This generates the resulting shifted RF signals, which are centered at 27.8 and 28.4 GHz. We report the linearization performance of the SO solution with: 1) the unshifted ICM; 2) the original ICM (i.e., estimated from the shifted signal); and 3) the ICM generated with the proposed method, presented in Section III-B. The performance of the GN solution is also presented as a reference. As can be seen from the obtained results in Fig. 5, the SO with the unshifted ICM does not converge to any solution since the ICM is not able to describe the statistics of the shifted signal. Second, the SO with the original ICM provides the optimal linearization performance, essentially equal to GN, since the ICM is directly calculated from the shifted signal. Third, the SO with the ICM estimated using the proposed method in Section III-B also provides optimal performance, despite the substantially lower associated complexity. The TRP-ACLR and the EVM
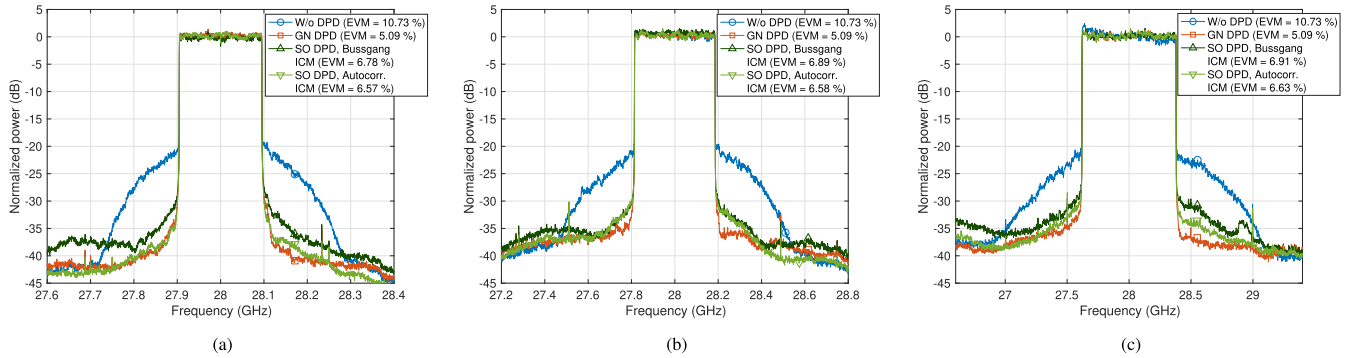
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

PASCUAL CAMPO *et al.*: INVERSE CM ESTIMATION FOR LOW-COMPLEXITY CLOSED-LOOP DPD SYSTEMS

13

Fig. 6. OTA linearization performance at EIRP $\approx$ +40.5 dBm for the SO method when the ICM is estimated through the two novel methods presented in Section IV, with the signal bandwidth of (a) 200, (b) 400, and (c) 800 MHz. The GN method is also measured and shown for reference.

numbers for the proposed method are in both cases maintained above +31 dBc and below 5.6%, respectively, thus satisfying the 3GPP standards [7] at FR-2. This performance is equal to that of the GN method, which involves heavy complexity in each DPD iteration. These results verify the effectiveness of the proposed ICM shifting method in terms of linearization performance.

E. Estimating ICM From the Autocorrelation Function

This fourth experiment validates the two novel methods presented in Section IV, which estimates the ICM from the autocorrelation function of the input data. These techniques are tested with three different signal bandwidths of 200 MHz (120 kHz SCS, 132 RBs), 400 MHz (120 kHz SCS, 264 RBs), and 800 MHz (120 kHz, 528 RBs). In all cases, the ICM is estimated using the proposed approach described in Section IV and then injected in the SO learning rule. The GN method is also measured for reference, while all the remaining system parameters are maintained as explained before.

The obtained measurement results are presented in Fig. 6. With a bandwidth of 200 MHz, the SO DPD with autocorrelation ICM estimation achieves an excellent linearization performance, very close to that of the GN approach. The SO DPD with the Bussgang ICM estimation lies somewhat behind but still obtains a good amount of linearization. This difference is reduced when considering the wider bandwidths of 400 and 800 MHz, in which the performance of the proposed novel solutions is very close to each other and also to the reference GN model. It is noted that there is no direct theory-based reason why the performance gap between the autocorrelation-based ICM estimation and the Bussgang ICM estimation methods is largest in the case of 200 MHz bandwidth. The difference can be stemming, e.g., from a slight change in the hardware setup, or by a minor movement in the position of the person handling the measurements. In all cases, the SO solution in combination with the methods proposed in Section IV successfully satisfies the +28 dBc TRP-ACLR and the 8% EVM limits while showing a remarkable complexity reduction in estimating the ICM.

F. Convergence Analysis

We continue the experimental results with a convergence analysis of the proposed DPD methods in a static waveform scenario. For this analysis, we consider the signal bandwidth of 800 MHz (120-kHz SCS, 528 RBs, 625 ksamples) and present the convergence behavior of the SO method when: 1) ICM is estimated normally from the input signal; 2) ICM is estimated with the autocorrelation method presented in Section IV-A; and 3) ICM is estimated with the Bussgang-based method presented in Section IV-B. The convergence speeds of the remaining methods are the same as that of 1), and thus, they are not shown separately. The remaining DPD parameters are maintained as stated above.

The convergence behavior of the DPD models is presented in Fig. 7, in terms of the TRP-ACLR and the NMSE. As can be seen from the figure, the TRP-ACLR convergence of the GN, the SO with classical ICM calculation, and the SO with autocorrelation-based ICM estimation is very similar, reaching the steady-state performance in around 15 CL iterations, after which the DPD behavior stabilizes at around +32 dBc TRP-ACLR. The convergence speed of the SO with the Bussgang ICM estimation is slightly slower, reaching the steady state in around 20 block iterations. However, the final values of TRP-ACLR are similar compared to the previous methods. This behavior is expected since the Bussgang-based method provides only an approximate ICM. Similar conclusions can be drawn from the second figure, which shows that the NMSE obtained with the different methods. These results show the favorable performance of the proposed SO solutions, which together with the achieved complexity reductions demonstrates very appealing performance–complexity tradeoffs.

G. DPD Learning Complexity With Different ICM Methods

In this final subsection, we gather the quantitative DPD learning complexity numbers obtained with the proposed ICM estimation methods and corresponding to the DPD parameterization of $P = 9$, $M = 4$, $C = 20$, $l = 5$, and $K = 20\,000$. Here, we measure the complexity in terms of average number of real multiplications per linearized sample, assuming that the DPD learning algorithms are executed over 15 closed-loop iterations—which is generally the number of iterations that the
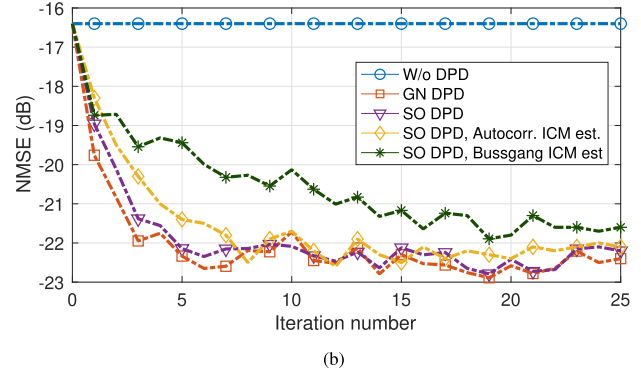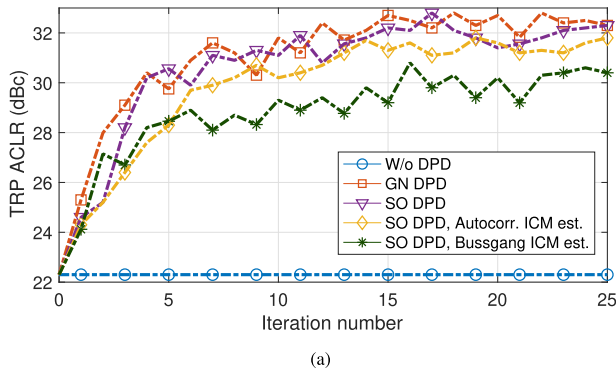
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

14   IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES

Fig. 7.   OTA convergence performance with 800 MHz waveform bandwidth and EIRP $\approx$ +40.5 dBm for the SO method and its variants, in terms of (a) TRP-based ACLR and (b) NMSE. The GN method is also measured and shown for reference.

TABLE II

AVERAGE LEARNING PATH COMPLEXITIES OF THE PROPOSED METHODS PER LINEARIZED SAMPLE AND THE RELATIVE COMPLEXITY REDUCTIONS WITH RESPECT TO THE REFERENCE GN APPROACH

DPD method	Average learning comp. (real mul. per lin. sample)	Percentage red. wrt. GN
Reference GN	1,681	0 %
SO + classical ICM calc.	186.85	88.9 %
SO + ICM subset	80.10	95.2 %
SO + freq. shift ICM	80.08	95.2 %
SO + autocorr. est. ICM	81.54	95.1 %
SO + Bussgang's est. ICM	81.42	95.1 %

models need to converge, as demonstrated in the earlier subsections. The obtained quantitative complexity results, building on the expressions in Table I, together with the corresponding relative complexity reduction with respect to the reference GN method, are presented in Table II.

As can be seen from the table, excellent complexity savings can be obtained through the proposed solutions. In all cases, the numerical average complexity is drastically reduced when comparing with the reference GN method. This is also reflected in the achieved percentage reduction, which is in most cases larger than 90%. Keeping in mind that all the proposed methods satisfy the 3GPP transmit waveform quality specifications [7], as shown through the obtained results in Section VI-B–E, the combination of efficient linearization performance and low processing and learning complexity is paving the way toward fast and continuous DPD adaptation with on-chip real-time implementations in commercial systems.

VII. CONCLUSION

In this article, various methods to efficiently estimate the inverse of the input data CM were proposed and combined with SO closed-loop learning in a DPD-based linearization context. The inherent low complexity of the SO learning combined with the proposed methods, allowed for remarkably reduced complexities in the DPD system, while maintaining a similar linearization performance compared to state-of-the-art solutions. To substantiate this, complexity analyses dealing

with the proposed solutions were performed, and thorough RF measurement results at 28 GHz mmW band, featuring a 64-element active antenna array, were presented. The obtained results, both in terms of performance and complexity, indicated very favorable performance–complexity tradeoffs of the proposed methods when comparing against the state of the art. The proposed methods are thus promising candidates for linearizing mmW phased-array transmitters as well as cellular terminal transmitters, the processing complexity being a key concern in both applications.

REFERENCES

[1] F. M. Ghannouchi and O. Hammi, "Behavioral modeling and predistortion," *IEEE Microw. Mag.*, vol. 10, no. 7, pp. 52–64, Dec. 2009.

[2] Y. Rahmatallah and S. Mohan, "Peak-to-average power ratio reduction in OFDM systems: A survey and taxonomy," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 1567–1592, 4th Quart., 2013.

[3] 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.

[4] S. Afsardoost, T. Eriksson, and C. Fager, "Digital predistortion using a vector-switched model," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 4, pp. 1166–1174, Apr. 2012.

[5] 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, Nov. 2001.

[6] A. Zhu, "Digital predistortion and its combination with crest factor reduction," in *Digital Front-End in Wireless Communications and Broadcasting: Circuits and Signal Processing*, F.-L. Luo, Ed. Cambridge, U.K.: Cambridge Univ. Press, 2011, ch. 9, pp. 244–279.

[7] *NR: Base Station (BS) Conformance Testing, Part 2*, document 3GPP TS 38.141-2, Version 15.1.0, Release 15, Mar. 2019.

[8] 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.

[9] C. Mollén, E. G. Larsson, U. Gustavsson, T. Eriksson, and R. W. Heath, Jr., "Out-of-band radiation from large antenna arrays," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 196–203, Apr. 2018.

[10] L. Anttila, A. Brihuega, and M. Valkama, "On antenna array out-of-band emissions," *IEEE Wireless Commun. Lett.*, vol. 8, no. 6, pp. 1653–1656, Dec. 2019.

[11] M. Abdelaziz, L. Anttila, A. Brihuega, F. Tufvesson, and M. Valkama, "Digital predistortion for hybrid MIMO transmitters," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 445–454, Jun. 2018.

[12] X. Liu *et al.*, "Beam-oriented digital predistortion for 5G massive MIMO hybrid beamforming transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 7, pp. 3419–3432, Jul. 2018.

[13] A. Brihuega *et al.*, "Piecewise digital predistortion for mmWave active antenna arrays: Algorithms and measurements," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 9, pp. 4000–4017, Sep. 2020.

[14] H. Holma, A. Toskala, and T. Nakamura, *5G Technology: 3GPP New Radio*, 1st ed. Hoboken, NJ, USA: Wiley, 2019.

[15] A. S. Tehrani, H. Cao, T. Eriksson, M. Isaksson, and C. Fager, "A comparative analysis of the complexity/accuracy tradeoff in power amplifier behavioral models," *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 6, pp. 1510–1520, Jun. 2010.

[16] F. Mkadem, A. Islam, and S. Boumaiza, "Multi-band complexity-reduced generalized-memory-polynomial power-amplifier digital predistortion," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 6, pp. 1763–1774, Jun. 2016.

[17] A. Zhu and T. J. Brazil, "An overview of Volterra series based behavioral modeling of RF/microwave power amplifiers," in *Proc. IEEE Annu. Wireless Microw. Technol. Conf.*, Dec. 2006, pp. 1–5.

[18] A. Zhu and T. J. Brazil, "Behavioral modeling of RF power amplifiers based on pruned Volterra series," *IEEE Microw. Wireless Compon. Lett.*, vol. 14, no. 12, pp. 563–565, Dec. 2004.

[19] H. Wang, G. Li, C. Zhou, W. Tao, F. Liu, and A. Zhu, "1-bit observation for direct-learning-based digital predistortion of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 7, pp. 2465–2475, Jul. 2017.

[20] A. Zhu, M. Wren, and T. J. Brazil, "An efficient Volterra-based behavioral model for wideband RF power amplifiers," in *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 2, Jun. 2003, pp. 787–790.

[21] M. Abdelaziz, L. Anttila, A. Kiayani, and M. Valkama, "Decorrelation-based concurrent digital predistortion with a single feedback path," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 1, pp. 280–293, Jan. 2018.

[22] N. Guan, N. Wu, and H. Wang, "Model identification for digital predistortion of power amplifier with signed regressor algorithm," *IEEE Microw. Wireless Compon. Lett.*, vol. 28, no. 10, pp. 921–923, Oct. 2018.

[23] P. P. Campo, V. Lampu, A. Meirhaeghe, J. Boutellier, L. Anttila, and M. Valkama, "Digital predistortion for 5G small cell: GPU implementation and RF measurements," *J. Signal Process. Syst.*, vol. 92, pp. 475–486, May 2020.

[24] L. Ding, R. Raich, and G. T. Zhou, "A Hammerstein predistortion linearization design based on the indirect learning architecture," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 3, May 2002, pp. 2689–2692.

[25] P. P. Campo, L. Anttila, D. Korpi, and M. Valkama, "Cascaded spline-based models for complex nonlinear systems: Methods and applications," *IEEE Trans. Signal Process.*, vol. 69, pp. 370–384, 2020.

[26] X. Hong, Y. Gong, and S. Chen, "A Wiener model for memory high power amplifiers using B-spline function approximation," in *Proc. 17th Int. Conf. Digit. Signal Process. (DSP)*, Jul. 2011, pp. 1–5.

[27] 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.

[28] P. P. Campo *et al.*, "Gradient-adaptive spline-interpolated LUT methods for low-complexity digital predistortion," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 68, no, 1, pp. 336–349, Jan. 2021.

[29] A. Molina, K. Rajamani, and K. Azadet, "Digital predistortion using lookup tables with linear interpolation and extrapolation: Direct least squares coefficient adaptation," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 3, pp. 980–987, Mar. 2017.

[30] Y. Ma, Y. Yamao, Y. Akaiwa, and C. Yu, "FPGA implementation of adaptive digital predistorter with fast convergence rate and low complexity for multi-channel transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 11, pp. 3961–3973, Nov. 2013.

[31] F.-L. Luo, *Digital Front-End in Wireless Communications and Broadcasting: Circuits and Signal Processing*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[32] L. Ding and G. T. Zhou, "Effects of even-order nonlinear terms on power amplifier modeling and predistortion linearization," *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 156–162, Jan. 2004.

[33] H. Enzinger, K. Freiberger, G. Kubin, and C. Vogel, "Baseband Volterra filters with even-order terms: Theoretical foundation and practical implications," in *Proc. 50th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2016, pp. 220–224.

[34] N. Lashkarian, J. Shi, and M. Forbes, "A direct learning adaptive scheme for power-amplifier linearization based on Wirtinger calculus," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 61, no. 12, pp. 3496–3505, Dec. 2014.

[35] A. Zhu, J. C. Pedro, and T. J. Brazil, "Dynamic deviation reduction-based Volterra behavioral modeling of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 54, no. 12, pp. 4323–4332, Dec. 2006.

[36] L. Guan and A. Zhu, "Simplified dynamic deviation reduction-based Volterra model for Doherty power amplifiers," in *Proc. Workshop Integr. Nonlinear Microw. Millim.-Wave Circuits*, Vienna, Austria, Apr. 2011, pp. 1–4.

[37] P. Jardin and G. Baudoin, "Filter lookup table method for power amplifier linearization," *IEEE Trans. Veh. Technol.*, vol. 56, no. 3, pp. 1076–1087, May 2007.

[38] P. P. Campo *et al.*, "Closed-loop sign algorithms for low-complexity digital predistortion: Methods and performance," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 1, pp. 1048–1062, Jan. 2021.

[39] Q. Luo, X. Zhu, C. Yu, and W. Hong, "Single-receiver over-the-air digital predistortion for massive MIMO transmitters with antenna crosstalk," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 1, pp. 301–315, Jan. 2020.

[40] X. Wang, Y. Li, C. Yu, W. Hong, and A. Zhu, "Digital predistortion of 5G massive MIMO wireless transmitters based on indirect identification of power amplifier behavior with OTA tests," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 1, pp. 316–328, Jan. 2020.

[41] X. Liu, W. Chen, L. Chen, F. M. Ghannouchi, and Z. Feng, "Linearization for hybrid beamforming array utilizing embedded over-the-air diversity feedbacks," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 12, pp. 5235–5248, Dec. 2019.

[42] G. H. Golub and C. F. Van Loan, *Matrix Computations*, vol. 3. Baltimore, MD, USA: JHU Press, 2012, pp. 238–239.

[43] I. Reed, "On a moment theorem for complex Gaussian processes," *IRE Trans. Inf. Theory*, vol. 8, no. 3, pp. 194–195, Apr. 1962.

[44] T. Araujo and R. Dinis, "On the accuracy of the Gaussian approximation for the evaluation of nonlinear effects in OFDM signals," *IEEE Trans. Commun.*, vol. 60, no. 2, pp. 346–351, Feb. 2012.

[45] J. Minkoff, "The role of AM-to-PM conversion in memoryless nonlinear systems," *IEEE Trans. Commun.*, vol. COM-33, no. 2, pp. 139–144, Feb. 1985.

[46] I. Cosme, I. Fernandes, J. Carvalho, and S. X.-D. Souza, "Block recursive matrix inverse," Dept. Engenharia Computaçao Automaçao, Univ. Federal Rio Grande Norte, Natal, Brazil, Tech. Rep., 2016.

[47] C. F. Van Loan, "The ubiquitous Kronecker product," *J. Comput. Appl. Math.*, vol. 123, nos. 1–2, pp. 85–100, 2000.

[48] W.-H. Steeb and Y. Hardy, *Matrix Calculus and Kronecker Product: A Practical Approach to Linear and Multilinear Algebra*, 2nd ed. Singapore: World Scientific, 2011.

[49] L. Guan and A. Zhu, "Low-cost FPGA implementation of Volterra series-based digital predistorter for RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 4, pp. 866–872, Apr. 2010.

[50] S. S. Haykin, *Adaptive Filter Theory*. London, U.K.: Pearson, 2008, pp. 226–227.

Pablo Pascual Campo (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees in telecommunications and electrical engineering from the Universidad Politécnica de Madrid, Madrid, Spain, in 2016 and 2018, respectively. He is currently pursuing the D.Sc. degree at the Department of Electrical Engineering, Tampere University, Tampere, Finland.

His research interests include digital predistortion, full-duplex systems and applications, and signal processing for wireless communications at the mmWave bands.

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

Since 2016, he has been a University Researcher with the Department of Electrical Engineering, Tampere University. From 2016 to 2017, he was a Visiting Research Fellow with Aalto University, Finland. His research interests are in radio communications and signal processing, with a focus on the radio implementation challenges in systems such as 5G, full-duplex radio, and large-scale antenna systems.

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

16   IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES

Vesa Lampu received the B.Sc. and M.Sc. degrees in electrical engineering from Tampere University, Tampere, Finland, in 2017 and 2019, respectively, where he is currently pursuing the D.Sc. degree.

He has been with the Department of Electrical Engineering, Tampere University, since 2018. His research focuses on energy-efficient multiple-input–multiple-output systems. He has also been involved in research on nonlinear system identification, with applications in full-duplex and digital predistortion techniques.

Neng Wang received the Ph.D. degree in electrical engineering from Queen's University, Kingston, ON, Canada, in 2005.

In 2006, he was with the Communications Research Centre, Ottawa, ON, Canada, as a Research Scientist, worked on signal processing for wireless communications. From 2007 to 2009, he was with Nortel Networks, Richardson, TX, USA, as a Member of Technical Staff, worked on 3GPP LTE and 3GPP2 standardization. From 2010 to 2016, he was with Qualcomm Corp R&D, Beijing, China, as a Senior Staff Engineer/Manager, worked on LTE modem algorithms, 4G Evolution, and 5G PHY/MAC systems design and standardization. In 2016, he joined the Huawei Research Center, Lund, Sweden, as an Algorithm Expert/Lead, working on 5G modem algorithms. His research interests include general areas of wireless communications PHY/MAC systems and signal processing.

Markus Allén received the B.Sc., M.Sc., and D.Sc. degrees in communications engineering from Tampere University of Technology, Tampere, Finland, in 2008, 2010, and 2015, respectively.

He is currently with the Department of Electrical Engineering, Tampere University, as a University Instructor. His current research interests include software-defined radios, 5G-related RF measurements, and digital signal processing for radio transceiver linearization.

Yan Guo received the B.E. degree in information science and engineering from East China Jiaotong University, Nanchang, Jiangxi, China, in 2007, the M.E. degree in communication and information systems from Southeast University, Nanjing, China, in 2011, and the Ph.D. degree in electronic engineering from University College Dublin (UCD), Dublin, Ireland, in 2016.

He is currently a Principal Research Engineer and a Research Team Leader with the Wireless Terminal Chipset Algorithm Department, Hisilicon, Huawei. His main research interests include nonlinear behavioral modeling and digital linearization techniques for RF front-end of broadband UE.

Mikko Valkama (Senior Member, IEEE) received the D.Sc. (Tech.) degree (Hons.) from the Tampere University of Technology (TUT), Tampere, Finland, in 2001.

In 2003, he was a Visiting Post-Doctoral Research Fellow with SDSU, San Diego, CA, USA. He is currently a Full Professor and the Department Head of electrical engineering with Tampere University (TAU), Tampere. His research interests include radio communications, radio localization, and radio-based sensing, with emphasis on 5G and beyond mobile radio networks.