A System Approach for Efficiency Enhancement and Linearization Technique of Dual-Input Doherty Power Amplifier

CHOUAIB KANTANA[1], MOUCHACINE BENOSMAN [2], RUI MA [2], AND YUJI KOMATSUZAKI[3]

(Regular Paper)

[1]CNRS, CNAM, ESIEE Paris, Univ Gustave Eiffel, F-77454 Marne-la-Vallée, France
[2]Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA 02139 USA
[3]Information R&D Research Center, Mitsubishi Electric Corporation, Tokyo 100-8310, Japan

CORRESPONDING AUTHOR: Mouchacine Benosman (e-mail: benosman@merl.com).

ABSTRACT   In this paper, we propose an efficient system approach to improve the power efficiency of dual-input Doherty power amplifier (DIDPA) with maintaining its linearity level. An auto-tuning process based on a hybrid heuristic search control (HHSC) is applied to optimally define DIDPA configuration by optimizing its free parameters. Digital predistortion (DPD) is then integrated to linearize DIDPA using an optimal reduced-complexity model based on the segmentation approach. An optimal pruning process of the free parameters, based on hill-climbing (HC) heuristics, is proposed to reduce the HHSC complexity in order to refine the optimal DIDPA configuration. The system approach has been approved by experimental results, in different scenarios, using an LTE 20 MHz signal with a PAPR of 8 dB. DPD linearization under optimal DIDPA configuration improved linearity using a low-complex model with only 30 coefficients, which exhibited an error vector magnitude (EVM) of 2.5% and an adjacent channel power ratio (ACPR) of -50 dB at an averaged output power of 34 dBm. By updating the cost function coefficients and pruning the free parameters, DIDPA exhibited an EVM of 3%, an ACPR of -50 dB, and a drain efficiency of 47% at an average output power of 39 dBm.

INDEX TERMS   Doherty power amplifier, dual-input, digital predistortion, control process, optimization, heuristics, pruning approach.

I. INTRODUCTION

Introducing non-constant amplitude modulated signal in radio frequency (RF) transmission systems enhances power amplifier (PA) efficiency while maintaining an adequate linearity level more challenging. Starting from wideband code division multiple access (W-CDMA) in 3G, the peak-to-average power ratio (PAPR) of signals increases with the use of orthogonal frequency division multiplex (OFDM) in 4G LTE and 5G new radio (NR).

Advanced PA architectures based on dynamic load or supply modulation have been proposed in the literature to avoid wasting excessive power resources [1]. Some of the most popular solutions are Doherty [2], envelope tracking [3], Chireix [4]. and outphasing [5]. These highly efficient topologies require linearization techniques such as digital predistortion (DPD) to meet the linearity requirement, especially with the increase of signal bandwidth.

The amplification architectures based on active load modulation are among the most common PA efficiency enhancement techniques (such as Doherty), which rely on the nonlinear interaction between the main and auxiliary transistors for modulated signals with a large dynamic. Although these architectures can be designed with a single RF input to be used in the transmitter, several studies have been reported in the literature to highlight the benefits of maintaining separate inputs [6], [7], [8] and the advantages of dual-input PA compared to single-input has been studied in [9].

The separation of RF inputs provides additional degrees of freedom, so-called free parameters, that can be set to

improve the performance or to enhance the PA efficiency [10]. These free parameters present a set of crucial circuit and system-level parameters, including PA bias voltages, power ratio, and phase shift of the separate input signals.

Focusing on dual-input Doherty PA (DIDPA), searching for optimal free parameters guaranteeing high performance requires experimental cross-validation or exhaustive search. These processes are usually costly and computationally significant, especially when the search space is enormous and not limited. Besides, setting these free parameters to their optimal values within a defined interval can be considered a global optimization problem. Several techniques have been proposed in the literature to find the optimal set of free parameters among large tunable ranges considered as search intervals [11].

Generally speaking, DIDPA should operate at its maximum efficiency, leading to poor linearity. To overcome this system drawback, it is necessary to introduce an efficient system approach to ensure the trade-off between linearity and efficiency. Regarding linearity, digital predistortion (DPD) is a powerful linearization technique used to compensate for the PA nonlinearities [12].

DPD consists of applying a pre-correction to the input signal so that the cascaded system (DPD and PA) behaves like an ideal linear and memoryless amplification. Designing a DPD model that allows finding the inverse characteristic of DIDPA is challenging, particularly finding a friendly candidate model for the hardware implementation with good numerical properties and high modeling accuracy.

On the other hand, the efficiency of DIDPA is limited by an output back-off (OBO) needed to prevent signal peaks from going beyond the saturation point. Since DPD introduces a back-off that allows the OBO to be equal to the PAPR of the input signal, efficiency can be further improved by reducing the PAPR signal with various crest factor reduction (CFR) techniques [13]. In the literature, several research works have focused on the joint combination of CFR and DPD [14], [15], [16], [17].

The motivation of this paper is to provide a rigorous and practical system approach to DIDPA. This is achieved by placing DIDPA in an iterative system-based process while focusing on three aspects related to the study of DIDPA: optimizing the free parameters provided by the separation of the input signal, reducing the PAPR using CFR, and compensating for the nonlinearities using DPD linearization.

A. STATE-OF-THE-ART
In the literature, the design of DIDPA with enhanced efficiency has been reported in numerous research works. Few of them deal with the optimization of its configuration, and even fewer deal with the joint optimization of DIDPA parameters and its linearization technique.

Early work investigating the linearization of DIDPA has been reported in [18], where the authors used vector-switched generalized memory polynomials [19] to improve the linearity. The separation of the RF inputs has been statically achieved by performing several combinations in a simulation environment.

In [20], the authors have split the RF inputs for multi-input Doherty PA by performing an exhaustive search that returns the best configuration regarding the power ratio. However, the bias voltages of Doherty PA are identically biased, which may limit the efficiency enhancement.

Another research work in [21] proposed an adaptive signal separation in terms of the power ratio from the static measured results of DIDPA. The results in [21] once again confirmed the interest in optimizing the separation of the input signals to be transmitted to DIDPA, which can significantly improve efficiency.

Another confirmation can be found in [22], where the authors highlight the need for optimizing the phase shift between the two RF inputs for dual-input load modulated balanced PA. In [22], the authors have performed an exhaustive search to find the optimal phase shift by sweeping the phase over a determined interval. This approach can be costly, especially in terms of its implementation. Besides, sweeping the phase over a large interval with reduced resolution can be critical, especially at intervals where there will be no output power, leading to heat dissipation in the device, which can damage it. Finding the optimal shift phase can be viewed as an optimization problem with a global minimum that requires a unidirectional minimization with reduced complexity.

The first work related to the online learning-based optimization of DIDPA is proposed in [23]. The authors proposed an adaptive technique based on a simultaneously perturbed stochastic approximation (SPSA) algorithm to tune the power ratio, the phase shift, and the bias voltages as free parameters. A cost function is used to control the algorithm convergence and defined in terms of the output gain G and the power added efficiency (PAE) in the additive criterion. However, the linearity requirement has not been met since SPSA has focused only on efficiency enhancement.

An extension of the work in [23] is reported in [24], where the authors update the cost function by including, in addition to G and PAE, the output power P_{out} and adjacent channel power ratio (ACPR) as metric referring to the linearity. The optimization process used in [24] is carried out by using a global optimization algorithm such as simulated annealing (SA) in the first order. Once the cost function achieves its optimal value, a fine-tuning process based on an approach of learning-based control such as on the extremum-seeking control (ESC) is used. In [24], the free parameters' global optimization process does not include any linearization technique or PAPR reducer.

In [25], a novel auto-tuning approach has been proposed to enhance the power efficiency of the DIDPA while meeting the linearity requirement. It consists of optimizing the free parameters using a proposed hybrid heuristic search control (HHSC) according to a designed cost function that indicates the trade-off between power efficiency and linearity. The work presented in [25] is considering the first part of the research work presented in this paper.

B. MAIN CONTRIBUTION

In this paper, we propose a new system approach to follow the workflow of the efficiency enhancement of DIDPA with its linearization. Three major phases conduct this workflow: optimization of DIDPA configuration based on HHSC, DPD linearization, and joint optimization or pruning version of HHSC and DPD. These three phases use an adaptive cost function and change its coefficients at each phase according to each specification requirement. The free parameters optimized in the HHSC process cover the baseband calibration process, such as the power level and PAPR reduction. Besides, a DPD linearization is jointly optimized and integrated into our approach to meet the linearity requirements. The proposed approach brings a set of contributions regarding the optimization and linearization of DIDPA, which are emphasized by:

- An efficient optimization process based on HHSC by combining a global optimization problem and adaptive control to ensure the convergence of the free parameters to their optimum.
- Design a multi-objective optimization cost function to represent the trade-off between linearity and power efficiency.
- Design an optimal DPD model with a reduced-complexity aspect by hill-climbing (HC) heuristic. The DPD model involved in this study is based on the model with a segmentation approach.
- An adaptive tuning of the cost function according to the linearity improvement by the DPD and power efficiency by the optimal configuration.
- An optimal process to prune the free parameters in HHSC to refine the optimal configuration with the presence of DPD towards reduced computational complexity.

Unlike conventional solutions reported in the literature for DIDPA that focus on a single perspective study, the advantage of our proposed system approach is to drive DIDPA iteratively to an optimal operating point where linearity and efficiency meet their requirements. This is achieved by thoroughly studying and optimizing each phase of the system approach and executing the appropriate algorithms with high performance, low complexity, and good numerical properties for its implementation.

The remainder of this paper is organized as follows. Section II presents the system-level aspects, including the DIDPA, the proposed architecture, and the free parameters to be optimized. Section III describes the proposed system approach by focusing on the cost function design, the auto-tuning process-based architecture, the DPD linearization technique, and re-optimizing free parameters using an optimal pruning process. Section IV describes the test bench. Section V presents the experimental results of the proposed approach. Finally, the conclusion is given in Section VI.

II. SYSTEM-LEVEL ASPECTS

A. DUAL-INPUT DOHERTY POWER AMPLIFIER

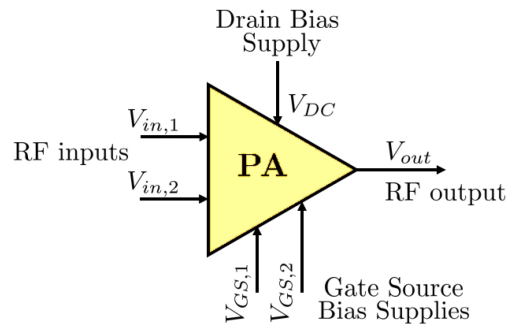The PA based on active load modulation such as Doherty and outphasing with separate RF inputs can be

FIGURE 1. Block diagram of dual-input PA.

viewed, by generalization, as the block diagram depicted in Fig. 1.

The DIDPA has two RF inputs, a drain bias V_{DC}, and two gate-source voltages $V_{GS,1}$ and $V_{GS,2}$ to control the transistor's terminal independently. A typical example of PA with independent V_{GS} is Doherty, where the main (carrier) amplifier is biased in class B and the auxiliary amplifier (typically named peaking amplifier) in class C.

The instantaneous amplitude and phase of each input in the baseband, as well as the V_{GS} gate bias voltages, can be controlled and adjusted separately, allowing a significant degree of freedom for these parameters to improve the power efficiency. The main amplifier reaches its maximum output voltage at a given operating back-off and becomes maximally efficient. From this power level, the auxiliary amplifier turns on and injects current into the common node, increasing the output power and modulating the load seen by the main amplifier. In this paper, the DIDPA presented in [26] is used as the device under test (DUT), where the authors have presented a 3.0–3.6−GHz wideband GaN Doherty PA with a frequency dependency compensating circuit.

B. ON-LINE ARCHITECTURE

To meet the objectives of optimizing DIDPA, which consists of improving efficiency while maintaining a better linearity level, we propose an on-line architecture described in Fig. 2.

In this paper, the DPD is used to linearize the DUT by compensating for the nonlinearities of DIDPA. Additionally, CFR is used to reduce the PAPR of the transmitted signal so that DIDPA can operate with less BO. Both CFR and DPD are implemented in baseband.

On the other hand, the DIDPA requires two separate input signals. Therefore, the baseband signal to be sent to the PA should be divided into two input signals, which are different in amplitude and phase, using a splitting function, the so-called digital splitter, designed in the baseband.

Each block has parameters to be set or controlled, which requires a design of a control engine based on an optimization approach that optimally determines these parameters to ensure an operating point of DUT exhibiting a better trade-off between efficiency and linearity.
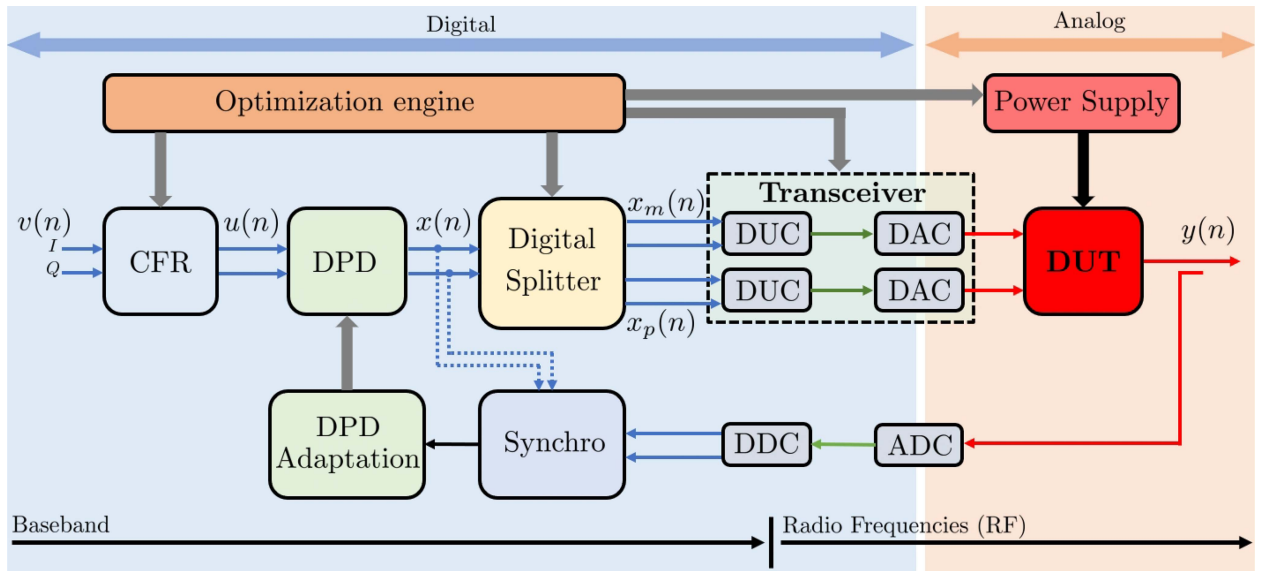
FIGURE 2. Block diagram of the on-line architecture.

TABLE 1. Free Parameters of the Proposed Architecture

Block	Free-parameter	Symbol	Unit
CFR	Threshold of PAPR reduction	μ	dB
Digital splitter	Power ratio	α	%
	Phase shift	ϕ	Degree
Transceiver	Attenuation difference	ψ	dB
DUT	Main bias voltage	$V_{GS,m}$	V
	Peaking bias voltage	$V_{GS,p}$	V

In this architecture, the DPD block will not be controlled by this control engine since the DPD technique requires linear regression techniques such as the least square (LS) method to identify the model coefficients.

C. FREE PARAMETERS
Each block in Fig. 2 has free parameters to be set or controlled, which are summarized in Table 1.

1) FREE PARAMETER OF CFR
The CFR technique used to reduce the PAPR is based on peak cancellation [27], [28], the principle of which is based on the clipping and filtering CFR technique and carried out through two stages: hard clip and clip-and-filter.

The hard clip is the most basic CFR technique, where the input signal $v(n)$ is clipped according to a threshold μ. The signal generated at the first step is expressed as:

$$v_{HC}(n) = \begin{cases} v(n) - v(n)\frac{\mu}{|v(n)|} & \text{if } |v(n)| > \mu \\ 0 & \text{if } |v(n)| \leq \mu \end{cases} \quad (1)$$

At the second stage, $v_{HC}(n)$ is filtered using noise shaping. Finally, the output signal $u(n)$ is given by subtracting a time-aligned weighted version of the filtered peak cancellation signal from the original input signal $v(n)$.

$$u(n) = v(n-d) - \alpha_s \times \text{filter}\{v_{HC}(n)\} \quad (2)$$

where α_s is the subtraction parameter.

Therefore, we use the clipping threshold μ as a free parameter of the CFR block to be controlled.

2) FREE PARAMETERS OF DIGITAL SPLITTER
The motivation behind using two separate RF inputs is to eliminate analog input splitters, such as the Wilkinson divider, and to allow independent power control to the main and peaking amplifier.

Digital splitter divides the complex signal $x = Xe^{j\theta}$ into two complex signals x_m and x_p defined as:

$$x_m = \alpha_m x \,;\; x_p = \alpha_p e^{-j\phi} x \quad (3)$$

We propose to take the power ratio α and the phase offset ϕ as free parameters with

$$\alpha_m = \sqrt{\alpha} \,;\; \alpha_p = \sqrt{1-\alpha} \quad (4)$$

3) FREE PARAMETER OF TRANSCEIVER
In the calibration process, it has been shown that two essential operations are needed to be established from the baseband: fixing the DAC resolution, which is integrated into the RF transceiver, and setting the gain attenuation, which controls the power level of the transmitted signal.

For DAC resolution, it is recommended to scale the IQ data in baseband to ensure high accuracy and minimize loss of information. The gain attenuation directly controls the power level of the signal in the Tx branch.

In the transceiver block, we use two parameters A_m and A_p defined in the baseband to control the attenuation in the branch T × 1 occupied by the main amplifier and in T × 2 occupied by the peaking amplifier.

We have approximately estimated the relationship between A_m (and A_p) and the average power of the main $P_{m,dBm}$ (and

peaking $P_{p,dBm}$) RF signal as:

$$P_{m,dBm}(A_m) = A_m + a\text{PAPR}(x_m) + b\sqrt{\alpha} + c$$

$$P_{p,dBm}(A_p) = A_p + a\text{PAPR}(x_p) + b\sqrt{1-\alpha} + c \quad (5)$$

where a, b, and c are parameters defined empirically from some preliminary tests and stored in look-up (LUT) and indexed in terms of the center frequency f_c, and the signal bandwidth. These test results provide a datasets of a, b, and c that will be used subsequently according to the parameters of the scenario at hands.

If A_m and A_p are equal, the power $P_{m,dBm}$ and $P_{p,dBm}$ should be equal. However, in practice, we have observed that by assigning the same numerical value to A_m and A_p, $P_{m,dBm}$ and $P_{p,dBm}$ through the power sensors are different. This difference between $P_{m,dBm}$ and $P_{p,dBm}$ can be adjusted and compensated in baseband by using a parameter denoted by ψ with

$$P_{m,dBm}(A_m) = P_{p,dBm}(A_p + \psi) \quad (6)$$

Finding ψ that satisfies (6) can be done in the calibration process. However, we propose to consider it as a free parameter controlled from baseband, which could be viewed as a hardware parameter since it can adjust the input power distribution over the main and peaking amplifiers.

4) FREE PARAMETER OF DUT

The main and peaking inputs of DIDPA control the main and peaking amplifiers, biased with $V_{GS,m}$ and $V_{GS,p}$, respectively. These biased voltages are controlled from the baseband and defined within a DC voltage range. The DC power supply used to manage the gate bias voltages is connected to the PC workstation through an Ethernet connection that enables real-time voltage monitoring from the baseband. Therefore, $V_{GS,m}$ and $V_{GS,p}$ are taken as free parameters.

III. SYSTEM APPROACH FOR LINEARIZATION AND EFFICIENCY ENHANCEMENT OF DIDPA

The process of the proposed system approach is mainly composed of five sub-processes:

- Design of the cost function to control the convergence of the free parameters optimization process.
- Optimization of free parameters based on the proposed auto-tuning process.
- DPD linearization based on ILA.
- Update the cost function designed in the first sub-process.
- Optimal pruning of free parameters in HHSC.

A. DESIGN OF COST FUNCTION

The cost function is an essential aspect in this research work to ensure a good trade-off between linearity and efficiency. The linearity requirement is presented in terms of two figures of merit (FOMs): error vector magnitude (EVM) and ACPR, while the efficiency requirement is presented by PAE and the output power P_{out}.

The EVM is a metric that measures the in-band distortion level. It is defined in the constellation domain and evaluates the deviation between the reference constellation point and the actual constellation point obtained in the presence of distortions. Analytically, EVM is defined as:

$$\text{EVM}_\% = \sqrt{\frac{\frac{1}{N}\sum_{j=0}^{N-1}(\delta I_j^2 + \delta Q_j^2)}{S_{avg}^2}} \times 100\% \quad (7)$$

Where δI and δQ are errors magnitude corresponding to in-phase symbol and quadrature symbol of received data compared with an ideally reconstructed constellation respectively, N is the number of symbols, S_{avg}^2 is the average square magnitude.

The ACPR is used to evaluate the out-band distortions and defined for the lower (left) and upper (right) adjacent channels as:

$$\text{ACPR}_{U,dB} = 10\log_{10}\frac{\int_{-B/2}^{B/2}P(y(t))df}{\int_{B/2}^{3B/2}P(y(t))df}$$

$$\text{ACPR}_{L,dB} = 10\log_{10}\frac{\int_{-B/2}^{B/2}P(y(t))df}{\int_{-3B/2}^{-B/2}P(y(t))df} \quad (8)$$

where B represents the bandwidth of the signal and $P(.)$ is power spectral density.

The FOMs are weighted according to their importance in the cost function. Additionally, some FOM penalization thresholds can also be defined when targeted specifications are not met.

In this paper, we propose to design the cost function J according to the weighted sum method [29] but with constraints [30] this ref. has not list of authors, which is defined as

$$J = w_1 J_{EVM} + w_2 J_{ACPR} + w_3 J_{PAE} + w_4 J_{P_{out}} \quad (9)$$

with

$$\begin{cases} J_{EVM} = w_1\left|\frac{EVM_t}{EVM}\right| \\ J_{ACPR} = \left|\frac{ACPR}{ACPR_t}\right| \\ J_{PAE} = \left|\frac{PAE}{PAE_t}\right| \\ J_{P_{out}} = \left|\frac{P_{out}}{P_{out,t}}\right| \end{cases} \quad (10)$$

where EVM_t, $ACPR_t$, PAE_t, and $P_{out,t}$ are EVM target, ACPR target, P_{out} target, and efficiency target, respectively, that the user attempts to reach.

The constraints of the cost function designed in (9) are defined as:

$$\begin{cases} \sum_{i=1}^{4} w_i = 1 \\ \max J_{FOM} = 1 \end{cases} \quad (11)$$

Since CFR as a nonlinear process deteriorates EVM dramatically, we propose to use in J_{EVM}, EVM of CFR denoted

by EVM_C, EVM of DPD and DUT denoted by EVM_{DD}, and EVM of the whole system including CFR, DPD, and DUT denoted by EVM_{CDD}.

There is a way to present the three EVMs in one feature, denoted by EVM_{ms} by using the mean square of EVM_{CDD}, EVM_C, and EVM_{DD}, which can be defined as:

$$EVM_{ms} = \sqrt{\frac{EVM_C^2 + EVM_{DD}^2 + EVM_{CDD}^2}{3}} \qquad (12)$$

The constraints of the cost function designed in (9) are defined as:

$$\begin{cases} w_1 = w_{1,1} + w_{1,2} + w_{1,3} \\ \sum_{i=1}^{4} w_i = 1 \\ \max J_{EVM} = \max J_{ACPR} = \max J_{PAE} = \max J_{P_{out}} = 1 \end{cases} \qquad (13)$$

The computation of ACPR is defined in terms of $ACPR_{L1}$ and $ACPR_{U1}$ as:

$$ACPR = 10 \log_{10} \left(\frac{10^{\frac{ACPR_{L1}}{10}} + 10^{\frac{ACPR_{U1}}{10}}}{2} \right) \qquad (14)$$

In (9), each objective function J_{FOM} is normalized by its target value $J_{FOM,t}$, which is defined as a user specification. The cost function design is carried out such that the optimization process attempts to maximize J to 1, indicating that the user's specifications are met.

B. HYBRID HEURISTIC SEARCH CONTROL

Finding the optimal configuration of each free parameter in Table 1 in a defined search range corresponds to our optimization problem.

The brute-force search can help find the optimal free parameters by exploring all possible combinations in the searching space. However, the brute-force search is not a practical solution to be implemented in real-time applications. Consequently, an auto-tuning approach based on an optimization algorithm is proposed to meet this need.

The proposed auto-tuning approach to optimize the free parameters is based on an efficient hybrid heuristic search control (HHSC) based on two types of model-free optimization methods: simulated annealing (SA) as a global optimization search and extremum-seeking control (ESC) as an adaptive control to fine-tune the optimized results.

The choice of combining SA and ESC stems from the main two properties of these optimization methods. Indeed, SA is well known to guarantee the convergence to a neighborhood of the global optimum in a compact search set [31]. Furthermore, ESC is proven in [32] to converge to a local optimum. Based on these two convergence properties, we choose to combine the SA, which will guide the free parameters to a neighborhood of the global optimum, and then switch to ESC, which will finite-tune the search for the optimal parameters in the local neighborhood of the optimum.

The vector of free parameters to be optimized is denoted by Θ. The cost function corresponding to Θ is denoted by $J(\Theta)$ or J for simplicity. The free parameters to be optimized are defined as:

$$\Theta = [\mu \; \alpha \; \phi \; \psi \; V_{GS,m} \; V_{GS,p}] \qquad (15)$$

The boundaries Θ_{min} and Θ_{max} are defined as well. The interval $[\Theta_{min} \; \Theta_{max}]$ presents the searching range of each free parameter. Some preliminary tests, or information about the system, especially DUT from previous works, are necessary to determine the proper optimization interval range.

1) SIMULATED ANNEALING

One of the best-known heuristic search methods for addressing the complex black-box global optimization problems is the SA algorithm proposed in [33].

Physical annealing in the metallurgy domain inspires the principle of the SA algorithm. Physical annealing is the process of heating a material until it reaches an annealing temperature. Then it will be cooled down slowly to increase the size of its crystals and reduce their defects. When the material is hot, the molecular structure is weaker and is more likely to change. When the material cools down, the molecular structure is more rigid and is less responsive to change.

Following the analogy with metallurgy, the slow cooling in simulated annealing depends on the slight decrease in the probability of accepting a worse solution as the solution space is explored. The algorithm should perform an extensive search to find the global optimum solution, so accepting worse solutions is fundamental.

SA algorithm is an almost straightforward stochastic search based on the Metropolis Monte Carlo method [34], the concept of which is to accepts not only the solutions that improve J, but also some solutions that worsen it with a probability p known as the Metropolis criterion and defined as:

$$p(\Delta E) = e^{-\frac{\Delta E}{k_{bolt} \mathcal{T}}} \qquad (16)$$

where ΔE is the change in cost function, k_{bolt} is Boltzmann's constant, and $\mathcal{T}$ is the control parameter analogous to the temperature of the annealing process.

During the search, the temperature is gradually decreased until reaching zero value in the perfect case.

The free parameters optimized using SA are denoted by $\Theta_{opt,SA}$. The cost function corresponding to $\Theta_{opt,SA}$ is denoted by $J_{opt,SA}$. The algorithm of SA is described in Algorithm (1).

2) EXTREMUM SEEKING CONTROL

Once SA algorithm has reached a neighborhood of $J_{opt,SA}$, the optimization procedure switches to ESC to fine-tune $\Theta_{opt,SA}$.

ESC is a control method that regulates the output of a dynamic map to its optimal value. Recently, it has been used extensively for real-time auto-tuning of physical systems, e.g. [35]. We propose here to use ESC in this context of real-time auto-tuning.

Indeed, ESC can be seen as a model-free optimization method, which does not need the gradient information explicitly, but estimates the gradient value of the cost function over

---

Algorithm 1: Algorithm of SA.

Initialization: $\mathcal{T}_0$, $\mathcal{T}_f$, $\mathcal{C}$, k_{bolt}, $iter = 0$
Define number of SA iterations $SA_{\max}$
Random initial solution Θ_0
$\Theta = \Theta_0$
$\Theta_{\text{opt,SA}} = \Theta$
Evaluate $J(\Theta)$
$\mathcal{T} = \mathcal{T}_0$
while $\mathcal{T} > \mathcal{T}_f$ **do**
 while $iter < SA_{max}$ **do**
  $iter = iter + 1$
  Select a random neighbor $\Theta' \in \Psi(\Theta)$
  $\Delta E = J(\Theta') - J(\Theta)$
  if $\Delta E > 0$ **then**
   $\Theta = \Theta'$
   if $J(\Theta')$ *is better than* $J(\Theta_{opt,SA})$ **then**
    $\Theta_{\text{opt,SA}} = \Theta'$
    $J_{\text{opt,SA}} = J(\Theta')$
   end
  else
   Generate uniformly distributed random
    $r \in [0\ 1]$
   if $r < e^{\frac{-\Delta E}{k_{bolt}\mathcal{T}}}$ **then**
    $\Theta = \Theta'$
   end
  end
 end
 $\mathcal{T} = \mathcal{C} \times \mathcal{T}$
 $iter = 0$
end
Return $\Theta_{\text{opt,SA}}$

Algorithm 2: Algorithm of HHSC.

Initialization: $\mathcal{T}_0$, $\mathcal{T}_f$, $iter = 0$, $\mathcal{K}$, a, ω
Run Algorithm of SA described in (1)
$\Theta_{0,\text{ESC}} = \Theta_{\text{opt,SA}}$
$\Theta = \Theta_{0,\text{ESC}}$
$J_{\text{old}} = J_{\text{opt,SA}}$
while *1* **do**
 Evaluate $J(\Theta)$
 if $J(\Theta) > J_{old}$ **then**
  $J_{\text{old}} = J(\Theta)$
  $\xi = J(\Theta) \times a\sin(\omega t)$
  $\hat{\Theta} = K \times \int \xi$
  $\Theta = \hat{\Theta} + a\sin(\omega t)$
 else
  $\Theta_{\text{opt,HHSC}} = \Theta$
  $J_{\text{opt,HHSC}} = J(\Theta_{\text{opt,HHSC}})$
  end while loop
 end
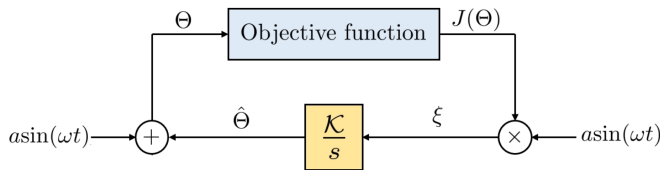**end**
Return $\Theta_{\text{opt,HHSC}}$

FIGURE 3. ESC scheme.

time, via properly designed feedback loops, i.e., filters. ESC is proven to converge to a local optima of the cost function.

One of the most popular and simple ESC methods is the perturbation-based ESC, which is proposed in [32], in which the concept is fundamentally depicted by the block diagram in Fig. 3, where it consists of a target system, the output of which is the cost function J, a perturbation signal $a\sin(\omega t)$, a gain $\mathcal{K}$, and an integrator.

According to Fig. 3, the controller $\frac{\mathcal{K}}{s}$ injects a perturbation signal $a\sin(\omega t)$ into the system, resulting in an output of the cost function $J(\Theta)$. This output is subsequently multiplied by $a\sin(\omega t)$, passed through the integrator $\frac{1}{s}$, leading to cost's gradient estimate Θ, and added to the perturbation signal $a\sin(\omega t)$.

The loop of ESC can be written as the following dynamical system

$$\xi = J(\Theta) \times a\sin(\omega t)$$

$$\frac{d\hat{\Theta}}{dt} = K \times \xi$$

$$\Theta = \hat{\Theta} + a\sin(\omega t) \tag{17}$$

It is worth noting that this ESC method only needs the numerical values of the cost J, and do not need any closed-from or numerical computations of the gradient of J. Its implementation is also rather straightforward, and corresponds to a simple forward Euler discretization of the dynamics given by (17). This simplicity of implementation makes ESC a good choice for online tuning of physical systems, when real-time computations capacities are limited or expensive, e.g., [35].

In our application, the perturbation-based ESC is used as an on-line process, which is placed downstream of SA to fine-tune $\Theta_{\text{opt,SA}}$, around the neighborhood of $\Theta_{\text{opt,SA}}$. The optimized configuration by ESC is denoted by $\Theta_{\text{opt,HHSC}}$. The cost function corresponding to $\Theta_{\text{opt,HHSC}}$ is denoted by $J_{\text{opt,HHSC}}$.

3) ALGORITHM OF HHSC
Starting from an initial solution

$$\Theta_0 = \begin{bmatrix} \mu_0\ \alpha_0\ \phi_0\ \psi_0\ V_{\text{GS,m}_0}\ V_{\text{GS,p}_0} \end{bmatrix}$$

SA optimizes the free parameters according to a designed cost function J. The optimized solution $\Theta_{\text{opt,SA}}$ returned by SA will be the initial solution for the ESC process. The algorithm of HHSC is described in Algorithm (2).

The optimized configuration by HHSC is denoted by $\Theta_{opt,HHSC}$, and its corresponding cost function is denoted by $J_{opt,HHSC}$.

C. DPD LINEARIZATION

Once HHSC is done, it will be decided whether to include DPD in the system approach or not. If not included, the cost function will be re-designed only based only on the efficiency requirement. However, the linearity specifications is more challenging to meet for the wideband signal scenarios. Hence, DPD linearization is part of system approach.

The decomposed vector rotation (DVR) model, proposed in [36], has been chosen to serve as a DPD model as it demonstrates its ability to linearize strong nonlinear behavior with memory [37].

The DVR model can be expressed by:

$$y(n) = \sum_{i=0}^{M_{lin}} a_i x(n-i) + \sum_{T_t \in S} T_t \tag{18}$$

where $x(n)$ and $y(n)$ are the input and output of the model, M_{lin} is the memory depth for the linear term, a_i are the complex coefficients of the linear term, and S is the set of the terms T_t which are used in the model with $S \subset \mathbf{T}$ and $\mathbf{T} = [T_{1,[0,...,P]}, T_2, T_3, T_4, T_5, T_6, T_7]$ where $T_{1,[0,...,P]}$ is the set $T_{1,0}, ..., T_{1,P}$ whose elements are defined hereafter along with $T_2, ..., T_7$.

$$T_{1,p} = \sum_{k=1}^{K} \sum_{i=0}^{M} c_{ki,1p} ||x(n-i)| - \beta_k| e^{j\theta(n-i)} \cdot |x(n)|^p$$

$$T_2 = \sum_{k=1}^{K} \sum_{i=0}^{M} c_{ki,2} ||x(n-i)| - \beta_k| \cdot x(n)$$

$$T_3 = \sum_{k=1}^{K} \sum_{i=0}^{M} c_{ki,3} ||x(n-i)| - \beta_k| \cdot x(n-i)$$

$$T_4 = \sum_{k=1}^{K} \sum_{i=0}^{M} c_{ki,4} ||x(n)| - \beta_k| \cdot x(n-i)$$

$$T_5 = \sum_{k=1}^{K} \sum_{i=0}^{M} c_{ki,5} ||x(n-i)| - \beta_k| \cdot x^2(n) \cdot x^*(n-i)$$

$$T_6 = \sum_{k=1}^{K} \sum_{i=0}^{M} c_{ki,6} ||x(n-i)| - \beta_k| \cdot x(n) \cdot |x(n-i)|^2$$

$$T_7 = \sum_{k=1}^{K} \sum_{i=0}^{M} c_{ki,7} ||x(n-i)| - \beta_k| \cdot x^*(n) \cdot x^2(n-i) \tag{19}$$

where K is the number of segments, β_k the bounds of the segments, p is the nonlinearity order, M the memory depth, and c_{ki} are the complex coefficients of the model for each segment.

The DVR model has been sized to find its optimal structure using the HC algorithm proposed in [38].
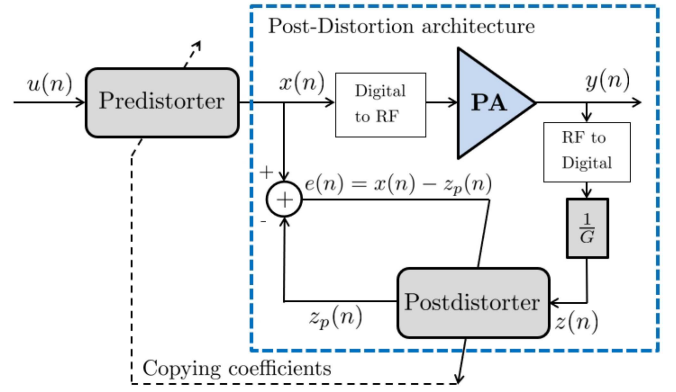

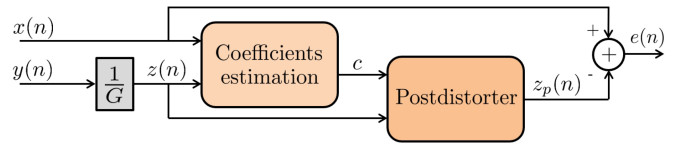FIGURE 4. Indirect learning architecture.

FIGURE 5. Principle of Post-Distortion architecture.

The indirect learning architecture (ILA), presented in Fig. 4, is used to identify the DPD coefficients, which are iteratively estimated by using LS approach in order to minimize the LS criterion built on the difference between the PA input $x(n)$ and $z_p(n)$, the model output, so-called the postdistorter, that is computed using the estimated model coefficient $\mathbf{c}$ and $z(n)$, the PA output divided by the amplification gain G.

Only the input-output signals from DIDPA are required in ILA to estimate the model coefficients. The principle is based on Post-Distortion and illustrated in Fig. 5.

The instantaneous error is defined as:

$$e(n) = x(n) - z_p(n) \tag{20}$$

The postdistorter input and output can be rewritten for N samples as

$$z_p = \mathbf{Z}\mathbf{c} \tag{21}$$

where $z_p = [z_p(1), ..., z_p(N)]^T$, $z = [z(1), ..., z(N)]^T$, $\mathbf{c}$ is a $C \times 1$ vector containing the set of estimated model coefficients, $\mathbf{Z}$ is $N \times C$ matrix of regressors containing basis functions of z. The LS solution will be the solution for the following equation for $\mathbf{c}$:

$$[\mathbf{Z}^H \mathbf{Z}] \mathbf{c} = \mathbf{Z}^H x \tag{22}$$

or equivalently using the pseudo-inverse:

$$\hat{\mathbf{c}} = [\mathbf{Z}^H \mathbf{Z}]^{-1} \mathbf{Z}^H x \tag{23}$$

where (22) minimizes the LS criterion:

$$\hat{\mathbf{c}} = \min_{\mathbf{c}}(J) \tag{24}$$

with:

$$J = \sum_{n=1}^{N} |e(n)|^2 = \sum_{n=1}^{N} |x(n) - z_p(n)|^2$$

D. UPDATE OF WEIGHTING COEFFICIENT

The DPD may apply a back-off to the operating point of DIDPA, which dramatically reduces the efficiency requirement. Therefore, the cost function J with DPD, which is denoted by $J_{\text{HHSC+DPD}}$, with updated FOMs and the same initial weighting coefficients will probably have deteriorated, and some free parameters in $\Theta_{\text{opt,HHSC}}$ will no longer be optimal.

To maximize the cost function again, we propose to rely on the design of the cost function J by adapting its weighting coefficients $w = [w_1 \, w_2 \, w_3 \, w_4]$ according to the change effected by the DPD. An intuitive approach can be used by attribute an equal weight to each FOM according to the following equation:

$$w_i = \frac{1}{n} \tag{25}$$

where $i = 1, 2, \ldots, n$ and n is the number of objective functions that present FOMs. In our context, the cost function is designed by combining efficiency and linearity, as is shown in (9). On the other hand, DPD with an efficient optimal DPD model can significantly improve linearity, which leads to reducing the weight of linearity FOMs in (9). Hence, we propose to design an adaptive cost function, in which the weight coefficients w are adaptive according to the improvement of linearity and efficiency over each block. The weighting coefficients w are updated with respect to how much DPD improves linearity FOMs, e.g., EVM and ACPR, compared to before applying it.

Starting from initial weighting coefficients w, we apply HHSC to optimize the free-parameters Θ with $J_{\text{opt,HHSC}}$, then DPD to linearize the DUT under the optimized free-parameters.

It is required to re-compute the cost function $J_{\text{HHSC+DPD}}$ once the DPD is performed, and compare it to $J_{\text{opt,HHSC}}$. if $J_{\text{HHSC+DPD}} < J_{\text{opt,HHSC}}$, we propose to update the weighting coefficients $w_{1,2}$, $w_{1,3}$, w_2, w_3, and w_4. $w_{1,1}$ is not concerned since it depends on the CFR operation.

In this process, we are only focusing on $w_{1,2}$, $w_{1,3}$, and w_2 that refers to linearity FOMs improved by DPD. We calculate the ratio n_i of the difference before and after DPD improvement for each linearity FOM. The ratio n_i is expressed as:

$$
\mathbf{n_1} = \left| \frac{\text{EVM}_{\text{HHSC,DD}} - \text{EVM}_{\text{HHSC+DPD,DD}}}{\text{EVM}_t} \right|
$$
$$
\mathbf{n_2} = \left| \frac{\text{EVM}_{\text{HHSC,CDD}} - \text{EVM}_{\text{HHSC+DPD,CDD}}}{\text{EVM}_t} \right|
$$
$$
\mathbf{n_3} = \left| \frac{\text{ACPR}_{\text{HHSC}} - \text{ACPR}_{\text{HHSC+DPD}}}{\text{ACPR}_t} \right| \tag{26}
$$

Algorithm 3: The Update of Weighting Coefficient w.

Initialize weighting coefficients
$\quad w = [w_{1,1} \; w_{1,2} \; w_{1,3} \; w_2 \; w_3 \; w_4]$
Apply HHSC
Get $J_{\text{opt,HHSC}}$
Measure EVM_{C}, EVM_{CD}, EVM_{CDD}, ACPR_{L1}, and
$\quad \text{ACPR}_{U1}$ before DPD
Compute $\text{ACPR}=f(\text{ACPR}_{L1}, \text{ACPR}_{U1})$ as in (14)
$\text{EVM}_{\text{HHSC,DD}}=\text{EVM}_{\text{DD}}$
$\text{EVM}_{\text{HHSC,CDD}}=\text{EVM}_{\text{CDD}}$
$\text{ACPR}_{\text{HHSC}}=\text{ACPR}$
Apply DPD using ILA
Measure EVM_{C}, EVM_{CD}, EVM_{CDD}, ACPR_{L1}, and
$\quad \text{ACPR}_{U1}$ with DPD
Compute $\text{ACPR}=f(\text{ACPR}_{L1}, \text{ACPR}_{U1})$
$\text{EVM}_{\text{HHSC+DPD,DD}}=\text{EVM}_{\text{DD}}$
$\text{EVM}_{\text{HHSC+DPD,CDD}}=\text{EVM}_{\text{CDD}}$
$\text{ACPR}_{\text{HHSC+DPD}}=\text{ACPR}$
Compute $J_{\text{HHSC+DPD}}$
if $J_{\text{HHSC+DPD}} < J_{\text{opt,HHSC}}$ **then**
$\quad \mathbf{n_1} = \left| \frac{\text{EVM}_{\text{HHSC,DD}} - \text{EVM}_{\text{HHSC+DPD,DD}}}{\text{EVM}_t} \right|$
$\quad \mathbf{n_2} = \left| \frac{\text{EVM}_{\text{HHSC,CDD}} - \text{EVM}_{\text{HHSC+DPD,CDD}}}{\text{EVM}_t} \right|$
$\quad \mathbf{n_3} = \left| \frac{\text{ACPR}_{\text{HHSC}} - \text{ACPR}_{\text{HHSC+DPD}}}{\text{ACPR}_t} \right|$
$\quad w_{1,2} = w_{1,2} \times \left(1 - \frac{10}{n_1}\right)$
$\quad w_{1,3} = w_{1,3} \times \left(1 - \frac{10}{n_2}\right)$
$\quad w_1 = w_{1,1} + w_{1,2} + w_{1,3}$
$\quad w_2 = w_2 \times (1 - n_3)$
$\quad w_3 = \frac{1 - w_1 - w_2}{2}$
$\quad w_4 = w_3$
else
$\quad$ Finish
end

Next, we update the weighting coefficients of FOM linearity as:

$$
w_{1,2} = w_{1,2} \times \left(1 - \frac{10}{n_1}\right)
$$
$$
w_{1,3} = w_{1,3} \times \left(1 - \frac{10}{n_2}\right)
$$
$$
w_1 = w_{1,1} + w_{1,2} + w_{1,3}
$$
$$
w_2 = w_2 \times (1 - n_3) \tag{27}
$$

Once $w_{1,2}$, $w_{1,3}$, and w_2 are updated, we propose to assign an equal weight between w_3 and w_4 according to:

$$
w_3 + w_4 = 1 - w_{1,1} - w_{1,2} - w_{1,3} - w_2
$$
$$
w_4 = w_3 \tag{28}
$$

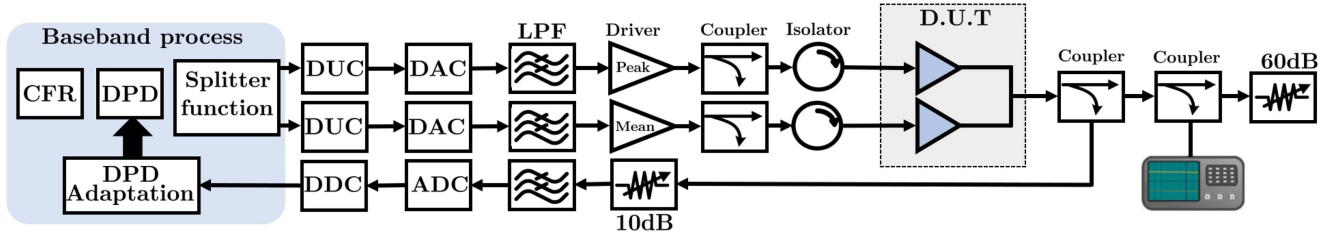The algorithm for updating the weighting coefficients w is described in Algorithm (3).

FIGURE 6. Block diagram of test bench.

E. OPTIMAL PRUNING OF FREE PARAMETERS IN HHSC

Once w is updated to the DPD contribution, the cost function's design J is changed.

We run the HHSC again, but only on reduced free parameters in Θ. The HHSC will only be performed on one free parameter that is the most sensitive one in Θ and has the most significant impact on the behavior of J. The pruning process is an off-line procedure, which aims to reduce the complexity of the HHSC when the weighting coefficients are updated.

Pruning the free parameters in HHSC is optimally achieved using the HC algorithm [39]. The motivation behind using the HC algorithm is that it is not a black-box optimization process. The neighborhood property in the HC algorithm makes it possible to follow the algorithm's evolution at each iteration.

Here, the cost function is used for the joint optimization of CFR and DUT.

The HC algorithm starts from a given initial element $\Theta_{1_{(0)},\text{HC}}$ at the first iteration and continually moves in the direction of the element with the best cost function value among its neighbors.

In the following, we denote $\Theta_{q_{(i)},\text{HC}}$ by $\Theta_{q_{(i)}}$, and $J_{q_{(i)},\text{HC}}$ by $J_{q_{(i)}}$.

At the qth iteration, the search procedure starts from $\Theta_{q_{(0)}}$ and test its neighbors $\Theta_{q_{(1)}}, \Theta_{q_{(2)}}, \ldots, \Theta_{q_{(M)}}$, where M is the number of neighbors of $\Theta_{q_{(0)}}$.

In this study, the neighborhood definition is inspired by the proposed one for the DVR model in [38]. The vector Θ consists of 6 free parameters: $[\mu \, \alpha \, \phi \, \psi \, V_{\text{GS},m} \, V_{\text{GS},p}]$. As these free parameters can have their values changed independently, they compose a 6-dimension space. The neighbor of element Θ is defined as an 6-tuple

$$\{\mu \times (1+\delta_\mu) \,; \, \alpha \times (1+\delta_\alpha) \,; \, \phi \times (1+\delta_\phi) \,;$$

$$\psi \times (1+\delta_\psi) \,; \, V_{\text{GS},m} \times (1+\delta_{V_{\text{GS},m}}) \,;$$

$$V_{GS,p} \times (1+\delta_{V_{GS,p}})\}$$

where $\delta_{\Theta(i)} \in [-1, 1] \times \frac{\Theta(i)}{10}$ with $i = 1, .., 6$.

The main property of this neighborhood definition is to apply the operation of $\delta_{\Theta(i)}$ to each free parameter $\Theta(i)$ individually.

According to this definition, the element

$$\Theta_{q_{(0)}} = \left[\mu \, \alpha \, \phi \, \psi \, V_{\text{GS},m} \, V_{\text{GS},p}\right]$$

at the qth iteration has 12 neighbors, which are:

$$\begin{cases}
\Theta_{q_{(1)}} = [\mu \times (1 + \frac{\mu}{10})\, \alpha \, \phi \, \psi \, V_{\text{GS,m}} \, V_{\text{GS,p}}] \\
\Theta_{q_{(2)}} = [\mu \times (1 - \frac{\mu}{10})\, \alpha \, \phi \, \psi \, V_{\text{GS,m}} \, V_{\text{GS,p}}] \\
\Theta_{q_{(3)}} = [\mu \, \alpha \times (1 + \frac{\alpha}{10})\, \phi \, \psi \, V_{\text{GS,m}} \, V_{\text{GS,p}}] \\
\Theta_{q_{(4)}} = [\mu \, \alpha \times (1 - \frac{\alpha}{10})\, \phi \, \psi \, V_{\text{GS,m}} \, V_{\text{GS,p}}] \\
\Theta_{q_{(5)}} = [\mu \, \alpha \, \phi \times (1 + \frac{\phi}{10})\, \psi \, V_{\text{GS,m}} \, V_{\text{GS,p}}] \\
\Theta_{q_{(6)}} = [\mu \, \alpha \, \phi \times (1 - \frac{\phi}{10})\, \psi \, V_{\text{GS,m}} \, V_{\text{GS,p}}] \\
\Theta_{q_{(7)}} = [\mu \, \alpha \, \phi \, \psi \times (1 + \frac{\psi}{10})\, V_{\text{GS,m}} \, V_{\text{GS,p}}] \\
\Theta_{q_{(8)}} = [\mu \, \alpha \, \phi \, \psi \times (1 - \frac{\psi}{10})\, V_{\text{GS,m}} \, V_{\text{GS,p}}] \\
\Theta_{q_{(9)}} = [\mu \, \alpha \, \phi \, \psi \, V_{\text{GS,m}} \times (1 + \frac{V_{\text{GS,m}}}{10})\, V_{\text{GS,p}}] \\
\Theta_{q_{(10)}} = [\mu \, \alpha \, \phi \, \psi \, V_{\text{GS,m}} \times (1 - \frac{V_{\text{GS,m}}}{10})\, V_{\text{GS,p}}] \\
\Theta_{q_{(11)}} = [\mu \, \alpha \, \phi \, \psi \, V_{\text{GS,m}} \, V_{\text{GS,p}} \times (1 + \frac{V_{\text{GS,p}}}{10})] \\
\Theta_{q_{(12)}} = [\mu \, \alpha \, \phi \, \psi \, V_{\text{GS,m}} \, V_{\text{GS,p}} \times (1 - \frac{V_{\text{GS,p}}}{10})]
\end{cases}$$

The element $\Theta_{q_{(i)}}$ with the maximized cost function $J_{q_{(i)}}$ is the solution denoted by $\Theta_{q_{(s)}}$. With the neighborhood definition, the best solution $\Theta_{q_{(s)}}$ can be compared with the initial solution $\Theta_{q_{(0)}}$ since only one free parameter is changed. An efficient way to do the comparison is to subtract $\Theta_{q_{(s)}}$ from $\Theta_{q_{(0)}}$, which make it easy to locate the position of the nonzero element in the vector $V = \Theta_{q_{(s)}} - \Theta_{q_{(s)}}$. The index of the nonzero element, denoted by idx, will then be stored in LUT, and the HC algorithm moves to the next iteration as long as $J_{q_{(s)}}$ is better than $J_{q-1_{(s)}}$. Otherwise, the HC algorithm stops.

The size of LUT is 6×2, where 6 refers to the number of free parameters in Θ (6 inputs). The cell corresponding to each free parameter, noted by pos, is incremented when the HC algorithm finds idx. Once the HC algorithm is finished, the free parameter to be used in HHSC, noted by Θ', is determined by the maximum incremented variable in the second column of LUT. If many free parameters have the same number of occurrences in LUT, the algorithm will take them as Θ'.

The optimal pruning of free parameters in HHSC is described in Algorithm 4.

The cost function that corresponds to the optimized free parameter Θ'_{opt}, denoted by $J_{\text{opt,upd}}$, will be compared to $J_{\text{HHSC+DPD}}$. If $J_{\text{opt,upd}}$ is better than $J_{\text{HHSC+DPD}}$, HHSC has improved FOMs compared to those from the previous optimal configuration, and a DPD is required to linearize the

FIGURE 7. Photo of test bench.

Algorithm 4: Optimal Pruning of Free Parameters.

Define cost function J
Choose the initial element $\Theta_0 = \Theta_{opt,HHSC}$
Set $q = 1$
$\Theta_{q_{(0)},HC} = \Theta_{0,HC}$
$J_{q_{(0)}} = J_{0,HC}$
while *(1)* **do**
    Define number of neighbors M of $\Theta_{q_{(0)}}$
    for $i \leftarrow M$ **do**
        Determine the neighbor $\Theta_{q_{(i)}}$
        Evaluate $J_{q_{(i)}}$
    end
    $\Theta_{q_{(s)}} = \text{argmin}_{\Theta_{q_{(i)}}} \left(J_{q_{(i)}} \right)$
    if $J_{q-1_{(s)}} < J_{q_{(s)}}$ **then**
        $V = \Theta_{q_{(s)}} - \Theta_{q_{(0)}}$
        Find index idx of nonzero element in V
        $\text{LUT}(q) = idx$
        $q = q + 1$
        $\Theta_{q_{(0)}} = \Theta_{q-1_{(s)}}$
    else
        end while loop
    end
end
$pos = $ Most frequent values in LUT
$\Theta' = \Theta(\max(pos))$
Optimize $J(\Theta')$ using J with updated w

DUT. Otherwise, the user must check whether the linearity-efficiency specifications are met or not.

IV. TEST BENCH

In order to validate the effectiveness of the proposed system approach, experiments have been carried out using a test bench. The block diagram and the photo of the test bench are shown in Figs. 6 and 7, respectively.

The DIDPA is controlled and evaluated using a MATLAB-based linearization and efficiency enhancement technique. The baseband IQ data are generated and split into two different IQ data inputs sent to the DUT through the AD9371 dual-channel RF transceiver, which is connected to Xilinx FPGA ZC700 through FMC connectors. The transceiver AD9371 up-converts the baseband signals to the carrier frequency f_c at 3 GHz. A Keysight N9010A MXA spectrum analyzer is used to characterize the signals spectrum at the output of DUT. For the observation path, The RF output signal is down-converted to the baseband by AD9371, which provides the baseband signal to the PC workstation. Around 100000 IQ samples were recorded for the baseband process with a sampling rate of 245.76 MHz.

The proposed system approach for DIDPA is tested using a 64-QAM modulated 20 MHz bandwidth LTE signal with a roll-off factor of 0.6 at $f_c = 3$ GHz with 8 dB of PAPR.
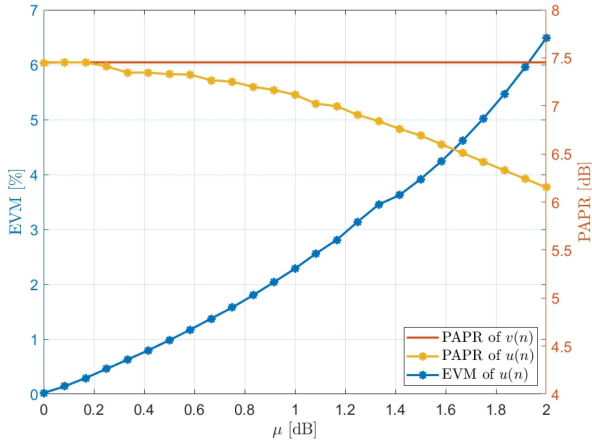
V. EXPERIMENTAL RESULTS
A. INITIAL CONDITIONS OF FREE PARAMETERS
We define the search range of the free parameters in Table 1 by setting the upper and lower bounds for each free parameter.

The search range is determined empirically according to some preliminary tests.

- Threshold of PAPR reduction μ: The CFR applies a non-linear process by clipping the input signal $v(n)$ according to a clipping threshold μ, which causes an EVM_C degradation. Fig. 8 shows the behavior of EVM_C and PAPR of the output $u(n)$ of the CFR block according to the variation of μ. PAPR of $v(n)$ decreases with respect to μ. The EVM of $v(n)$ degrades exponentially with increasing μ. As we target an EVM_t around 3%, we define the search interval $[0 \; \mu_{max}] = [0 \; 1.2]$, for which when $\mu = 1.2$, the degradation of the EVM by CFR is almost 3%.
- Power ratio α: Since α presents the power distribution between the main and peaking branches, we set its search range to $[0 \; 1]$.

FIGURE 8. Behavior of EVM_C and PAPR of $v(n)$ versus μ.

FIGURE 9. Evolution of cost function J over HHSC iterations.

TABLE 2. Optimal Free Parameters

	μ	α	ϕ	ψ	$V_{GS,m}$	$V_{GS,p}$	J
Θ_0	0	0.5	$40°$	0	-1.5	-2.5	0.57
$\Theta_{opt,SA}$	0.62	0.14	$164°$	1.54	-1.37	-1.95	0.65
$\Theta_{opt,HHSC}$	0.65	0.15	$164°$	1.53	-1.39	-1.95	0.66

- Phase shift ϕ: The search range $[-180° \quad 180°]$ is the standard interval to be set. However, based on some preliminary tests, it was shown that there is some critical search range to be avoided. Therefore, $[-10° \quad 200°]$ is taken as a search range for the free parameter ϕ.
- Attenuation difference ψ: As ψ is categorized as a hardware free parameter that refers to the physical power difference between the main and peaking amplifier, we set its search range to $[\psi_{min} \; \psi_{max}] = [-2 \; 2]$ dB.
- Bias voltage V_{GS}: We set the bias voltage search range for both $V_{GS,m}$ and $V_{GS,p}$ to $[V_{GS,min} \; V_{GS,max}] = [-4 \; -1]$ V in order to provide a flexible variation between deep-class C condition that should enhance efficiency and a near-class B bias where linearity should be improved. The drain bias V_{DC} is 28 V.

B. INITIALIZATION OF COST FUNCTION

At this point, it is necessary to determine the initial weighting coefficients w and the target FOMs.

We initialize the weighting coefficients

$$w = [w_1 \; w_2 \; w_3 \; w_4]$$

by:

$$w_{1,1} = \frac{0.1}{3} ; \; w_{1,2} = \frac{0.1}{3}; \; w_{1,3} = \frac{0.1}{3}$$
$$w_2 = 0.1$$
$$w_3 = 0.4$$
$$w_4 = 0.4$$

We attribute more weights to the efficiency since DPD will be included to linearize the DUT. This refers to the fact that the linearity requirements are more relaxed, as it is easier to meet with DPD, unlike efficiency.

Regarding target FOMs, they are defined as:

$$\begin{cases} EVM_t = 3\% \\ ACPR_t = -50 \, \text{dB} \\ PAE_t = 60\% \\ P_{out,t} = 40 \, \text{dBm} \end{cases}$$

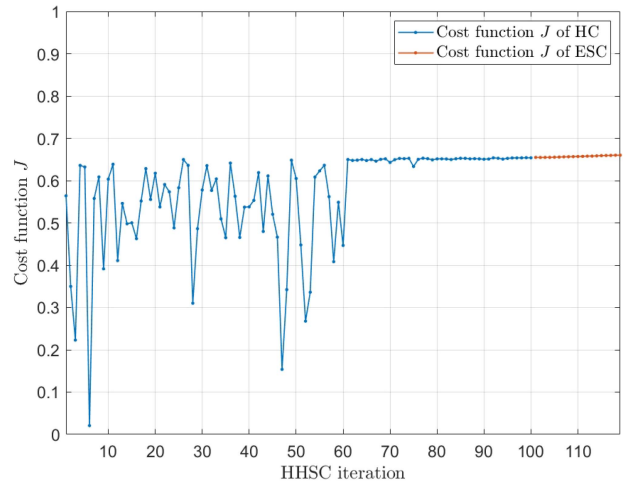### C. JOINT OPTIMIZATION OF CFR AND DUT

The initial solution for HHSC is defined as:

$$\Theta_0 = [\mu_0 \, \alpha_0 \, \phi_0 \, \psi_0 \, V_{GS,m_0} \, V_{GS,p_0}]$$
$$= [0 \, 0.5 \, 40° \, 0 \, -1.5 \, -2.5]$$

In this step, HHSC is performed according to Algorithm (2). For the SA algorithm, we set $\mathcal{T}_0 = 1$, $\mathcal{T}_f = 0.01$, $\mathcal{C} = 0.96$, $k_{bolt} = 1$, and $SA_{max} = 60$.

The convergence of J is shown in Fig. 9.

In HHSC, we set empirically 100 iterations for SA to converge, while ESC requires 20 iterations. As SA is defined as a stochastic optimization method, we can see from Fig. 9 that the stochastic behavior has been exhibited in the first 60 iterations where J evolves randomly.

After 61 iterations, the SA algorithm returns the optimal solution. After 120 iterations in total, HHSC returns the optimal configuration $\Theta_{opt,HHSC}$ summarized in Table 2.

The evolution of free parameters over HHSC is illustrated in Fig. 10, in which the free parameters have similar behavior to J according to the HHSC iterations.

The FOMs corresponding to $J_{opt,HHSC}$ are shown in Table 3.

It should be noted that HHSC shown in Fig. 10 was performed on the linear region of DIDPA. Indeed, since SA generates random solutions, it was decided to reduce the operating point of DUT while HHSC is running in order to set up a security measure of the DUT.

Once the free parameters are optimized, we raise the back-off by increasing the input power from 11.36 dBm to 21.16 dBm so that the system performs nearly at $P_{out,t}$.
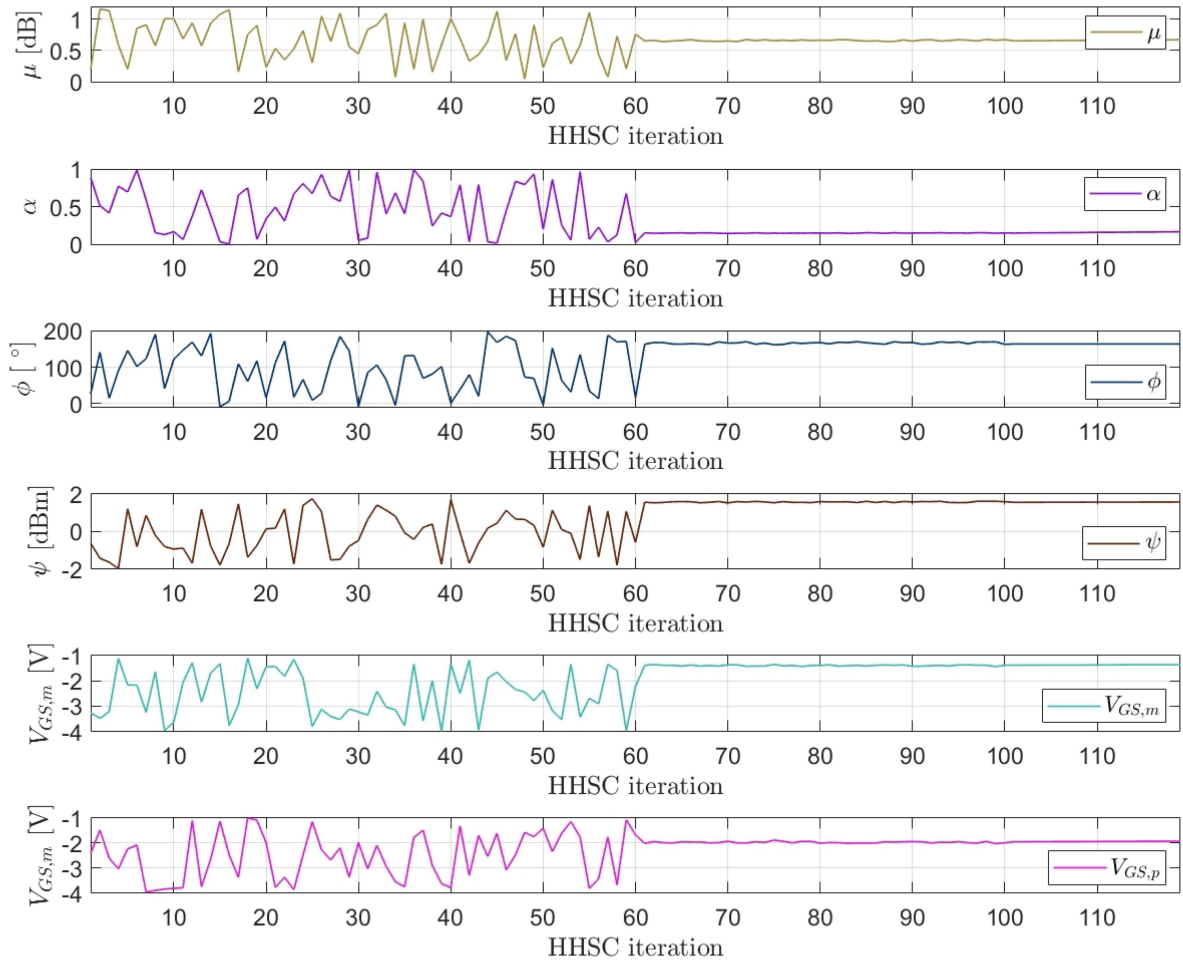
FIGURE 10. Evolution of free parameters over HHSC iterations.

TABLE 3. FOMs According to the Optimal Free Parameters $\Theta_{opt,HHSC}$ With $P_{in} = 11.36$ dBm

EVM_C [%]	1.48
EVM_{DD} [%]	2.34
EVM_{CDD} [%]	2.72
$ACPR_{U1}$ [dB]	-42.32
$ACPR_{L1}$ [dB]	-42.39
PAE [%]	16.21
P_{out} [dBm]	32.92
$J_{opt,HHSC}$	0.66

TABLE 4. FOMs According to the Optimal Free Parameters $\Theta_{opt,HHSC}$ With $P_{in} = 21.16$ dBm

EVM_C [%]	1.48
EVM_{DD} [%]	7.26
EVM_{CDD} [%]	8
$ACPR_{U1}$ [dB]	-27.44
$ACPR_{L1}$ [dB]	-27.15
PAE [%]	46.58
P_{out} [dBm]	35.63
$J_{opt,HHSC}$	0.81

The FOMs after increasing the input power are summarized in Table 4.

According to the results from Table 4, the efficiency FOMs are enhanced by sacrificing the linearity FOMs. This confirms the choice of the initial weighting coefficients where w_3 and w_4 have more influence than w_1 and w_2.
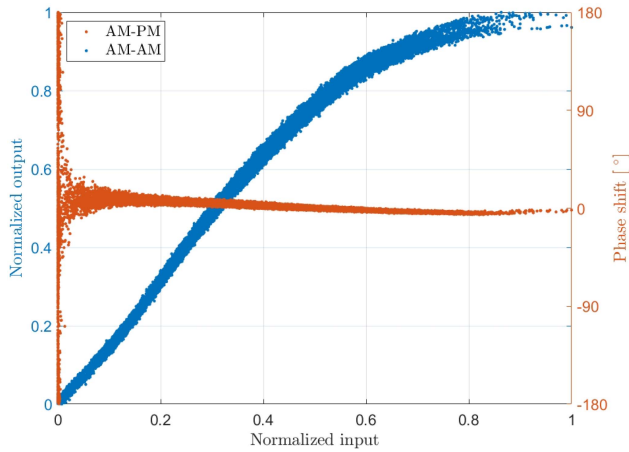
By assigning the free parameters to their optimal configuration $\Theta_{opt,HHSC}$ in Table 2, the DUT, including the digital splitter and transceiver, can eventually be seen as a single-input single-output system where the input is $x(n)$ and the output is denoted $y(n)$.

The AM-AM and AM-PM characteristics of DIDPA are shown in Fig. 11, where we can see a saturation at high power, which leads to strong nonlinearities. Besides, the memory effects are exhibited as well.

D. DPD LINEARIZATION

The DPD is carried out in two steps:
- Determination of the optimal DVR model.
- Convergence of linearity FOMs (ACPR and EVM) using ILA.

FIGURE 11. AM-AM and AM-PM curves.

FIGURE 12. Evolution of the HC algorithm in 3D in terms of *N*, *C* and *Cond* for sizing the DVR model.

1) DETERMINATION OF OPTIMAL DVR MODEL

According to the HC algorithm presented in [38], the structure of the DVR model is optimally sized.

In this study, the cost function, denoted by Y, is defined as a search criterion to ensure a good trade-off between three features: modeling accuracy presented by normalized mean square error (NMSE), denoted by N and computed between $x(n)$ and $z_p(n)$, model complexity presented by the number of coefficients C, and computational complexity, which is given by the condition number of the regressor matrix $\mathbf{Z^H Z}$ in (22), denoted by $Cond$.

The cost function Y is used to control the convergence of the HC algorithm in order to return an optimal DVR model with a considerable ability to be implemented in hardware. We define the cost function as:

$$Y = 0.5N + 0.25C + 0.25 \log_{10}(Cond) \qquad (29)$$

The hardware implementation relies on the numerical properties of the DPD model, which in this case, is presented by C and $Cond$. The design of the cost function in (29) may deteriorate C, and N since the objective function of $Cond$ is sized in such a way to override the influence of C and N slightly.

However, the deterioration of N can be overcome by the DPD convergence towards the solution that presents a better trade-off between linearization performance, complexity, and numerical properties.
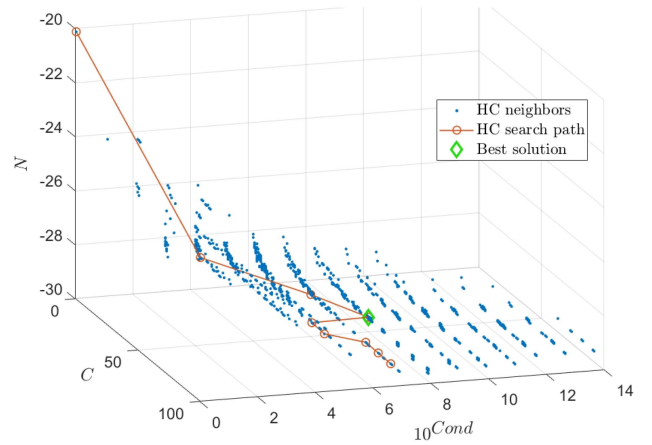
Fig. 12 illustrates the HC algorithm's evolution in terms of C, $Cond$, and N, where there are a total of 1659 DVR model structures tested through 9 HC iterations. The blue dots present the neighbors tested by the HC algorithm. The red dots indicate the search path taken by the HC algorithm. The green diamond highlights the best solution.

The parameters of the optimal DVR model structure are:

$$K = 4$$

$$\beta = [0.25\, 0.5\, 0.75]$$

$$M_{lin} = 1\,;\ M = 3$$

$$T_S = \begin{bmatrix} T_{1,0}\ T_{1,2}\ T_{1,4} \end{bmatrix}$$

with $C = 50$ coefficients, $10^{Cond} = 10^8$, and $N = -29.45$ dB.

Replacing the parameters of the optimal structure in (18), the optimal DVR model used in DPD is expressed by:

$$
\begin{aligned}
x(n) = & \sum_{i=0}^{1} a_i u(n-i) \\
& + \sum_{k=1}^{4} \sum_{i=0}^{3} c_{ki,1} ||u(n-i)| - \beta_k| e^{j\theta(n-i)} \\
& + \sum_{k=1}^{4} \sum_{i=0}^{3} c_{ki,2} ||u(n-i)| - \beta_k| e^{j\theta(n-i)} \cdot |u(n)| \\
& + \sum_{k=1}^{4} \sum_{i=0}^{3} c_{ki,3} ||u(n-i)| - \beta_k| e^{j\theta(n-i)} \cdot |u(n)|^4
\end{aligned}
$$

2) DPD USING ILA

As NMSE being considered as a strong indicator of the ILA convergence, Fig. 13 presents the evolution of the NMSE according to the DPD iterations. As can be seen, the convergence of NMSE is rapid, in which NMSE is improved significantly from the first DPD iteration. At the first DPD iteration, the coefficients of the DPD model are initialized by $c = [1, 0, \ldots, 0, 0]$, which make it a transparent block, by which $x(n) = u(n)$, and $z_p(n) = z(n)$.

EVM_{CDD} is improved from 8% to 2.97%, which is confirmed in Fig. 14 where the red dots present the IQ constellation of $y(n)$, and the blue dots are the reference IQ constellation of $v(n)$. ACPR has significantly been improved by over 20 dB, which can be confirmed in Fig. 15, where the output signal of DIDPA without DPD is shown in the red plot and with DPD in the green plot. Since the DVR model is optimally sized to ensure a good trade-off between performance linearization, model complexity, and numerical stability of the identification process. The numerical properties of the optimal
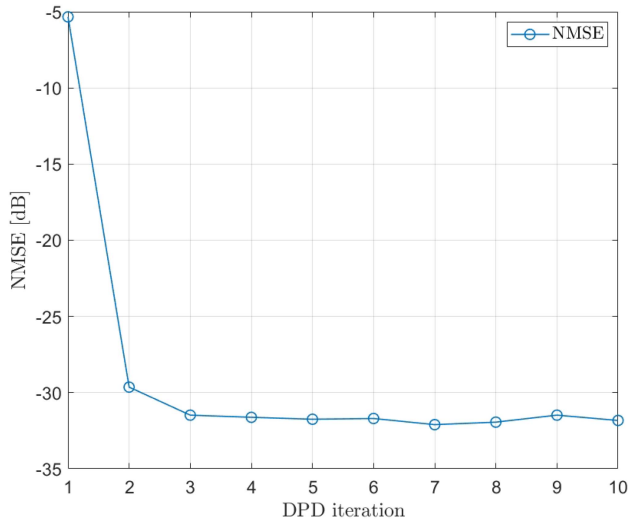
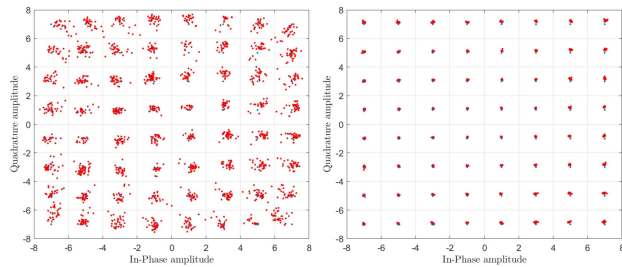**FIGURE 13. Evolution of NMSE according to DPD iterations.**
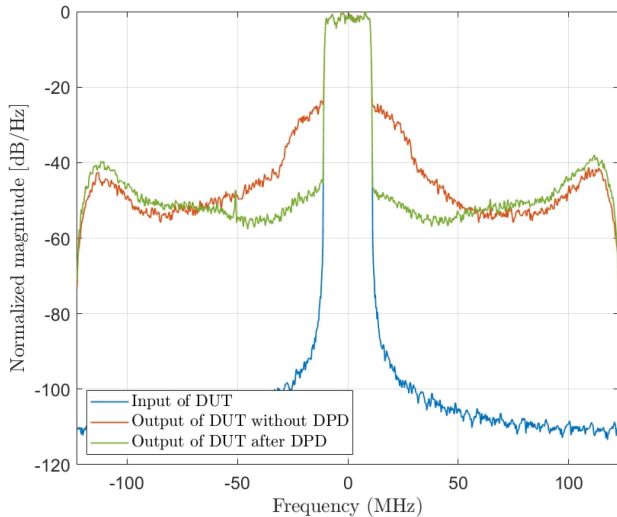

FIGURE 14. IQ constellation before and after DPD.

FIGURE 15. Spectra of the output signal of DUT without and with DPD.

DVR model at the final DPD iteration are:

$$C = 50$$

$$Cond = 10^8$$

$$\epsilon = 4$$

TABLE 5. FOMs According to the Optimal Free Parameters

	Before DPD	After DPD
EVM_C [%]	1.48	1.48
EVM_{DD} [%]	7.26	2.22
EVM_{CDD} [%]	8	2.97
$ACPR_{U1}$ [dB]	-27.44	-48.71
$ACPR_{L1}$ [dB]	-27.15	-47.86
PAE [%]	46.58	30.75
P_{out} [dBm]	35.63	34.09
$J_{opt,HHSC}$	0.81	0.78

where ϵ is the dynamic range of the model coefficients estimated by DPD.

Linearity FOMs are greatly improved by DPD, which closely meets the targeted linearity FOMs under the optimal configuration $\Theta_{opt,HHSC}$.

However, DPD applies a BO to DIDPA, which may deteriorate the efficiency. At the final DPD iteration, the efficiency FOMs are:

$$\begin{cases} PAE & = 30.75\% \\ P_{out} & = 34.09\,dBm \end{cases}$$

which does not meet the target FOMs ($P_{out,t}$ and PAE_t) set in the user specifications.

Table 5 summarizes the FOMs before and after DPD. The cost function is computed using FOMs before and after DPD.

From Table 5, the cost function is decreased after the DPD application. As DPD partially modifies the system conditions, the configuration $\Theta_{opt,HHSC}$ may not always be the optimal solution, which stimulus the motivation to update the weighting coefficients and re-launch a new HHSC.

E. UPDATE WEIGHTING COEFFICIENTS

The weighting coefficients w should be updated to take into account the linearity improvement by DPD to the DUT in which the efficiency is influenced.

According to Algorithm (3), the weighting coefficients w are updated as follows:

$$\begin{cases} w_{1,1} = 0.05\,;\ w_{1,2} = 0.01\,;\ w_{1,3} = 0.01 \\ w_2 = 0.05 \\ w_3 = 0.44 \\ w_4 = 0.44 \end{cases}$$

Being the weighting coefficients w reflect each FOM's impact in the value of the cost function J, Fig. 16 presents the weighting contribution of each FOM, of which it illustrates the contribution with the initial and updated weighting coefficients.

F. HHSC OPTIMAL PRUNING
1) HC ALGORITHM
Optimal pruning of free parameters in HHSC is performed with the presence of the DPD block where the predistorter
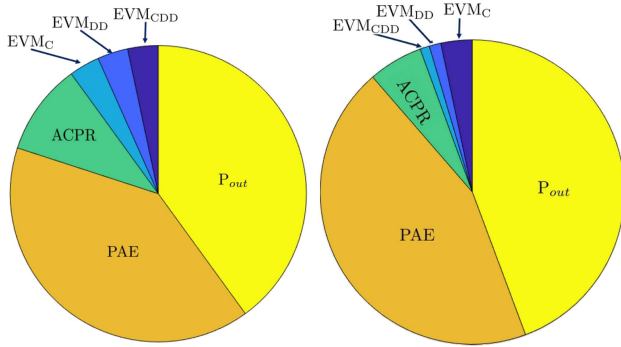
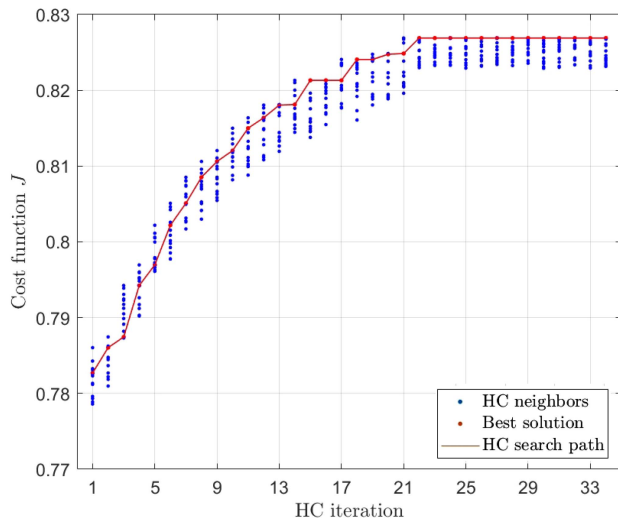**FIGURE 16. Impact distribution of FOMs in the design of *J* with initial and updated *w*.**

FIGURE 17. Optimal pruning of free parameters based on HC algorithm.
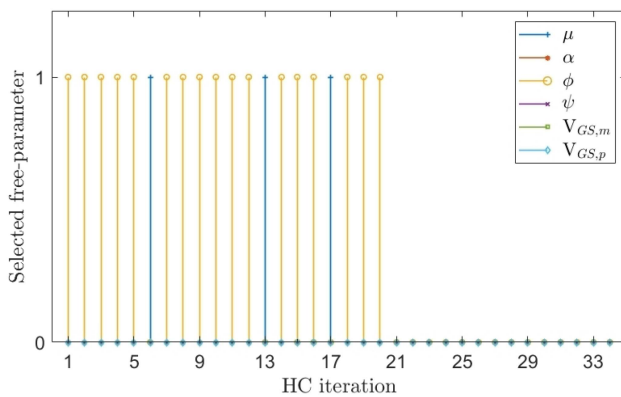

FIGURE 18. Selected free parameter over HC iterations.

is filled by the model coefficients identified at the final DPD iteration and presented.

Starting from the initial solution

$$\Theta_{0,\text{HC}} = \begin{bmatrix} 0.65\, 0.15\, 164^{\circ}\, 1.53\, -1.39\, -1.95 \end{bmatrix}$$
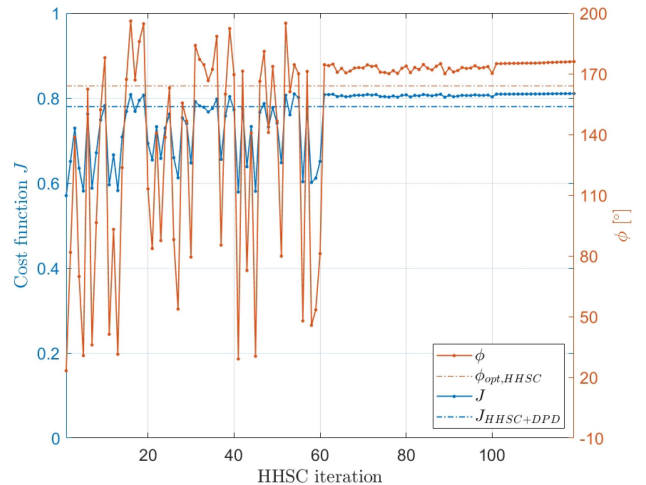

FIGURE 19. Evolution of cost function *J* and Θ′ over HHSC iterations.

the evolution of the HC algorithm to prune the free parameters for HHSC is illustrated in Fig. 17. The red dots present the solution at the HC iteration. The blue dots present the neighbors.

In order to investigate the behavior of the HC algorithm, Fig. 18 presents which free parameter has been stored in the dictionary through the HC iterations.

As can be seen, the free parameter ϕ has often been repeated, which means that the optimization behavior of HHSC is seen as most sensitive to ϕ.

On the other hand, the free parameter μ has been stored three times, where the configuration

$$\left[\mu \times \left(1 \pm \frac{\mu}{10} \right) \alpha\, \phi\, \psi\, V_{\text{GS,m}}\, V_{\text{GS,p}} \right]$$

has the best cost function J at the 6th, 13th, and 17th HC iteration.

After the 21th HC iteration, no configuration was found with a better J, which triggers the HC algorithm to stop.

2) HHSC WITH PRUNED FREE PARAMETERS

As discussed before, the goal of the HC algorithm here is to prune optimally the free parameters of HHSC, which aims to find the most sensitive free parameter to CFR and DUT with DPD.

Hence, $\Theta' = [\phi]$ is the most sensitive free parameter, which will be used in HHSC according to the cost function J designed with the updated weighting coefficients w.

Starting from the initial solution $\Theta'_0 = [164^{\circ}]$, which presents the optimal ϕ from the previous HHSC, Fig. 19 present the evolution of the cost function J and the free parameter ϕ according to the HHSC iterations.

According to Fig. 19, the free parameter ϕ has been re-optimized, where its optimal value becomes $\Theta'_{\text{opt}} = 175.9^{\circ}$ that corresponds to the optimal cost function $J_{\text{opt,upd}} = 0.81$. By comparing $J_{\text{opt,upd}}$ and $J_{\text{HHSC+DPD}}$, the DPD coefficients are required to be updated using ILA since the optimal configuration of DUT is changed.
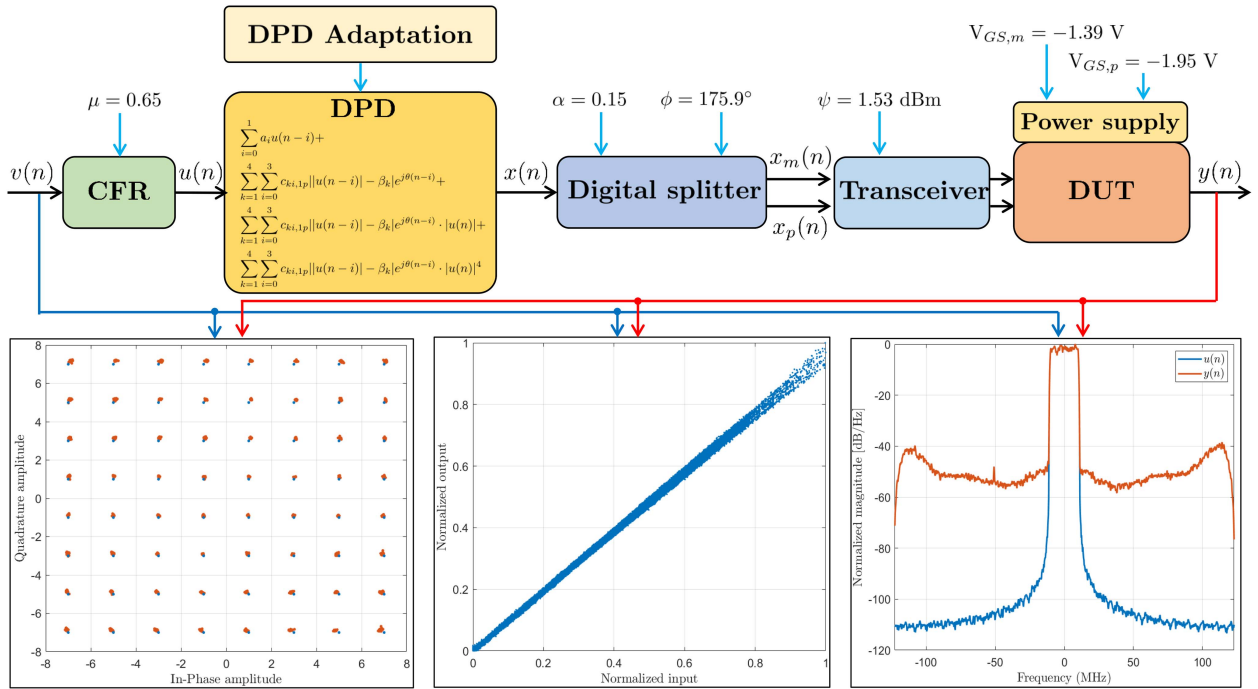
FIGURE 20. Final optimal configuration with DPD.

TABLE 6. FOMs According to the Optimal Free Parameters After the Optimal Pruning of Free Parameters in HHSC

	Before DPD	After DPD	After DPD with optimal pruning of free parameters
EVM_C [%]	1.48	1.48	1.48
EVM_{DD} [%]	7.26	2.22	2.04
EVM_{CDD} [%]	8	2.97	2.46
ACPR_{U1} [dB]	-27.44	-48.71	-49.21
ACPR_{L1} [dB]	-27.15	-47.86	-50.10
PAE [%]	46.58	30.75	36.11
P_{out} [dBm]	35.63	34.09	39.11
$J_{\text{opt,HHSC}}$	0.81	0.78	0.85

Table 6 summarizes the results, where FOMs are presented before and after applying DPD. At this level, the application of DPD consists only of updating the existing DPD coefficients in the predistorter since the optimal pruning of free parameters HHSC was performed with DPD.

With the new optimal configuration, DPD has improved the linearity FOMs. On the other hand, the efficiency FOMs are improved. This can be confirmed in Table 6, where the PAE is improved by almost 5.5%. By comparing the cost function, it can be seen that $J_{\text{HHSC+DPD}} = 0.85$ becomes better than $J_{\text{opt,HHSC}} = 0.82$, which leads to finishing the auto-tuning approach.

Fig. 20 illustrates the final architecture of DIDPA with CFR and DPD. The final optimal configuration is also highlighted. The AM-AM and AM-PM characteristics of the whole system (CFR+DPD+DUT) are shown in Fig. 20, along with the

spectra of the input-output signals and the IQ constellation of $u(n)$ plotted in blue and $y(n)$ plotted in red.

VI. CONCLUSION

In this paper, we proposed an auto-tuning approach to exploit at best dual-input Doherty PA to maximize power efficiency while being compliant with the linearity specifications. The proposed auto-tuning approach relies on conducting a global optimization combined with a control process to find the optimal configuration of a set of the crucial circuit and system-level parameters that are appropriately merged with the DPD linearization and the CFR technique. This proposed approach has been performed according to an adaptive designed cost function, representing the trade-off between efficiency and linearity. In order to sharpen the optimal configuration, we propose a new approach based on the HC algorithm to prune the free parameters optimally when DPD linearization is applied. The proposed approach has been validated through experimental results, in which we use a 20 MHz LTE signal scenario. The proposed approach to optimizing the dual-input Doherty PA has been well validated by presenting a good trade-off between linearity, computational complexity, and efficiency. Besides, the DPD model used, which is optimally sized, has very good numerical properties, making it a perfect candidate for its implementation on hardware such as FPGA.

REFERENCES

[1] D. J. Shepphard, J. Powell, and S. C. Cripps, "An efficient broadband reconfigurable power amplifier using active load modulation," *IEEE Microw. Wireless Compon. Lett.*, vol. 26, no. 6, pp. 443–445, Jun. 2016.

[2] W. H. Doherty, "A new high efficiency power amplifier for modulated waves," *Proc. Inst. Radio Engineers*, vol. 24, no. 9, pp. 1163–1182, Sep. 1936.

[3] F. Wang, A. H. Yang, D. F. Kimball, L. E. Larson, and P. M. Asbeck, "Design of wide-bandwidth envelope-tracking power amplifiers for OFDM applications," *IEEE Trans. Microw. Theory Techn.*, vol. 53, no. 4, pp. 1244–1255, Apr. 2005.

[4] T. Cappello, C. Florian, T. W. Barton, M. Litchfield, and Z. Popovic, "Multi-level supply-modulated chireix outphasing for LTE signals," in *Proc. IEEE MTT-S Int. Microw. Symp.*, 2017, pp. 1846–1849.

[5] H. Chireix, "High power outphasing modulation," *Proc. Inst. Radio Engineers*, vol. 23, no. 11, pp. 1370–1392, Nov. 1935.

[6] R. Darraji, F. M. Ghannouchi, and O. Hammi, "A dual-input digitally driven Doherty amplifier architecture for performance enhancement of Doherty transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 59, no. 5, pp. 1284–1293, May 2011.

[7] R. Darraji and F. M. Ghannouchi, "Digital Doherty amplifier with enhanced efficiency and extended range," *IEEE Trans. Microw. Theory Techn.*, vol. 59, no. 11, pp. 2898–2909, Nov. 2011.

[8] C. M. Andersson, D. Gustafsson, J. Chani Cahuana, R. Hellberg, and C. Fager, "A 1–3-GHz digitally controlled dual-RF input power-amplifier design based on a doherty-outphasing continuum analysis," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 10, pp. 3743–3752, Oct. 2013.

[9] A. Piacibello et al., "Comparison of S-band analog and dual-input digital doherty power amplifiers," in *Proc. IEEE 13th Eur. Microw. Integr. Circuits Conf.*, Madrid, 2018, pp. 269–272.

[10] R. Darraji, P. Mousavi, and F. M. Ghannouchi, "Doherty goes digital: Digitally enhanced Doherty power amplifiers," *IEEE Microw. Mag.*, vol. 17, no. 8, pp. 41–51, Aug. 2016.

[11] R. Horst, P. M. Pardalos, and N. Van Thoai, *Introduction to Global Optimization: Non-Convex Optimization and its Applications*. Berlin, Germany: Springer-Verlag, 2002.

[12] A. Katz, J. Wood, and D. Chokola, "The evolution of PA linearization: From classic feedforward and feedback through analog and digital predistortion," *IEEE Microw. Mag.*, vol. 17, no. 2, pp. 32–40, Feb. 2016.

[13] F. Sandoval, G. Poitau, and F. Gagnon, "Hybrid peak-to-average power ratio reduction techniques: Review and performance comparison," *IEEE Access*, vol. 5, pp. 27145–27161, 2017.

[14] R. N. Braithwaite, "A combined approach to digital predistortion and crest factor reduction for the linearization of an RF power amplifier," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 1, pp. 291–302, Jan. 2013.

[15] A. Mbaye, G. Baudoin, A. Gouba, Y. Louet, and M. Villegas, "Combining crest factor reduction and digital predistortion with automatic determination of the necessary crest factor reduction gain," in *Proc. IEEE 44th Eur. Microw. Conf.*, 2014, pp. 837–840.

[16] S. Wang, M. Roger, and C. Lelandais-Perrault, "Impacts of crest factor reduction and digital predistortion on linearity and power efficiency of power amplifiers," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 66, no. 3, pp. 407–411, Mar. 2019.

[17] S. Wang, M. Roger, J. Sarrazin, and C. Lelandais-Perrault, "A joint crest factor reduction and digital predistortion for power amplifiers linearization based on clipping-and-bank-filtering," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 7, pp. 2725–2733, Jul. 2020.

[18] J. C. Cahuana, P. Landin, D. Gustafsson, C. Fager, and T. Eriksson, "Linearization of dual-input Doherty power amplifiers," in *Proc. Int. Workshop Integr. Nonlinear Microw. Millimetre-Wave Circuits*, Leuven, 2014, pp. 1–3.

[19] Y. Liu, W. Pan, S. Shao, and Y. Tang, "A new digital predistortion for wideband power amplifiers with constrained feedback bandwidth," *IEEE Microw. Wireless Compon. Lett.*, vol. 23, no. 12, pp. 683–685, Dec. 2013.

[20] P. Singerl, T. Magesacher, and M. Mataln, "Predistortion- and development-platform for multi-input RF power amplifiers," in *Proc. IEEE 11th Eur. Microw. Integr. Circuits Conf.*, 2016, pp. 29–32.

[21] J. Peng, S. He, W. Shi, T. Yao, J. Wu, and J. Wang, "Adaptive signal separation for dual-input Doherty power amplifier," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 1, pp. 121–131, Jan. 2020.

[22] E. Guillena, W. Li, P. L. Gilabert, and G. Montoro, "Prediction of the optimal phase shift between control signals in dual-input power amplifiers," in *Proc. IEEE Int. Workshop Integr. Nonlinear Microw. Millimetre-Wave Circuits*, Cardiff, U.K., 2020, pp. 1–3.

[23] S. Niu, A. M. Koushik, R. Ma, K. H. Teo, S. Shinjo, and Y. Komatsuzaki, "Stochastically approximated multiobjective optimization of dual input digital Doherty power amplifier," in *Proc. IEEE 10th Int. Workshop Comput. Intell. Appl.*, Hiroshima, 2017, pp. 147–152.

[24] R. Ma et al., "Machine-learning based digital Doherty power amplifier," in *Proc. IEEE Int. Symp. Radio-Freq. Integration Technol.*, Melbourne, VIC, 2018, pp. 1–3.

[25] C. Kantana, R. Ma, M. Benosman, Y. Komatsuzaki, and K. Yamanaka, "A hybrid heuristic search control assisted optimization of dual-input Doherty power amplifier," in *Proc. IEEE 51st Eur. Microw. Conf.*, 2022, pp. 126–129.

[26] Y. Komatsuzaki, K. Nakatani, S. Shinjo, S. Miwa, R. Ma, and K. Yamanaka, "3.0–3.6 GHz wideband, over 46% average efficiency GaN Doherty power amplifier with frequency dependency compensating circuits," in *Proc. IEEE Topical Conf. RF/Microw. Power Amplifiers Radio Wireless Appl.*, Phoenix, AZ, 2017, pp. 22–24.

[27] W. J. Kim, K. J. Cho, S. P. Stapleton, and J. H. Kim, "Doherty feed-forward amplifier performance using a novel crest factor reduction technique," *IEEE Microw. Wireless Compon. Lett.*, vol. 17, no. 1, pp. 82–84, Jan. 2007.

[28] D. Lopez, P. L. Gilabert, G. Montoro, and N. Bartzoudis, "Peak cancellation and digital predistortion of high-order QAM wideband signals for next generation wireless backhaul equipment," in *Proc. Int. Workshop Integr. Nonlinear Microw. Millimetre-Wave Circuits*, Leuven, 2014, pp. 1–3.

[29] T. Murata, H. Ishibuchi, and M. Gen, "Specification of genetic search directions in cellular multi-objective genetic algorithms," in *Proc. Int. Conf. Evol. Multi-Criterion Optim.*, 2001, pp. 82–95.

[30] Y. Haimes "On a bicriterion formulation of the problems of integrated system identification and system optimization," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. SMC-1, no. 3, pp. 296–297, Jul. 1971.

[31] A. Lecchini-Visintini, J. Lygeros, and J. Maciejowski, "Simulated annealing: Rigorous finite-time guarantees for optimization on continuous domains," *Adv. Neural Inf. Process. Syst.*, 2007, pp. 865–872.

[32] K. B. Ariyur and M. Krstic, *Real Time Optimization by Extremum Seeking Control*. Hoboken, NJ, USA: Wiley Interscience, 2003.

[33] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[34] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, no. 6, 1953, Art. no. 1087.

[35] M. Benosman, *Learning-Based Adaptive Control: An Extremum Seeking Approach - Theory and Applications*. Oxford, U.K.: Butterworth-Heinemann, 2016.

[36] A. Zhu, "Decomposed vector rotation-based behavioral modeling for digital predistortion of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 2, pp. 737–744, Feb. 2015.

[37] C. Kantana, O. Venard, and G. Baudoin, "Comparison of GMP and DVR models," in *Proc. IEEE Int. Workshop Integr. Nonlinear Microw. Millimetre-Wave Circuits*, 2018, pp. 1–3.

[38] C. Kantana, O. Venard, and G. Baudoin, "Decomposed vector rotation model sizing by hill-climbing heuristic for digital predistortion of RF power amplifiers," in *Proc. IEEE Topical Conf. RF/Microw. Power Amplifiers Radio Wireless Appl.*, 2020, pp. 22–25.

[39] K. A. Sullivan and S. H. Jacobson, "A convergence analysis of generalized hill climbing algorithms," *IEEE Trans. Autom. Control*, vol. 46, no. 8, pp. 1288–1293, Aug. 2001.

CHOUAIB KANTANA received the Ph.D. degree from University Gustave Eiffel, Marne-la-Vallée, France, in 2021. He was a Visiting Research Intern with Mitsubishi Electric Research Laboratories, Cambridge, MA, USA. His research interests include wireless communications, digital predistortion, optimization problems, and deep learning for wireless communication systems.

MOUCHACINE BENOSMAN received the Ph.D. degree from Ecole Centrale de Nantes, Nantes, France, in 2002. He is currently a Senior Research Scientist with Mitsubishi Electric Research Labs (MERL), Cambridge, MA, USA. Before joining MERL, he was with Reims University, Reims, France, Strathclyde University, Glasgow, U.K., and the National University of Singapore, Singapore. His research interests include multi-agent distributed control with applications to robotics and smart-grids, control of infinite dimensional systems with applications to fluid dynamics, learning and adaptive control with applications to analog, and RF circuits auto-tuning. He is an Associate Editor for IEEE CONTROL SYSTEMS LETTERS, and *Journal of Optimization Theory and Applications*. He is a Senior Editor of *International Journal of Adaptive Control and Signal Processing*.

YUJI KOMATSUZAKI received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from Waseda University, Tokyo, Japan, in 2007, 2009 and 2012, respectively. Since 2012, he has been with the Information Technology Research and Development Center, Mitsubishi Electric Corporation, Kamakura, Japan, where he has been involved with the research and development of microwave amplifiers for telecommunication systems. From 2016 to 2017, he was a Visiting Scholar with the Center for Wireless Communication, University of California, San Diego, CA, USA.

RUI MA received the Dr.-Ing. degree from the University of Kassel, Kassel, Germany, in 2009. From 2010 to 2012, he was a Senior RF Research Engineer with Nokia Siemens Networks, Espoo, Finland. Since 2012, he has been with Mitsubishi Electric Research Labs, Cambridge, MA, USA, where he is a Senior Principal Scientist of RF Research, leading technologies development of advanced radio including applications of GaN RF power devices. From 2016 to 2021, he was a Visiting Scientist with THz Integrated Electronics Group, Massachusetts Institute of Technology, Cambridge, MA, USA. He holds more than 30 U.S. patents and patent applications. Dr. Ma is currently an Associate Editor for IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES. He was the recipient of the specification award by MIPI Alliance for the development of Analog Reference Interface for Envelope Tracking eTrak specification in 2013.