IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 8, 2019, accepted November 24, 2019, date of publication November 28, 2019,
date of current version December 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2956596

Theoretical Analysis of Nonlinear Amplification Effects in Massive MIMO Systems

SARA TEODORO[1], ADÃO SILVA[1], RUI DINIS[2], FILIPE M. BARRADAS[1], (Member, IEEE), PEDRO M. CABRAL[1], (Senior Member, IEEE), AND ATÍLIO GAMEIRO[1]

[1]DETI, Instituto de Telecomunicações, Universidade de Aveiro, 3810-193 Aveiro, Portugal
[2]Instituto de Telecomunicações, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Lisbon, Portugal

Corresponding author: Sara Teodoro (steodoro@av.it.pt)

This work was supported in part by the European Regional Development Fund (FEDER) through the Competitiveness and Internationalization Operational Program (COMPETE 2020) of the Portugal 2020 Framework, Regional OP Centro, under Grant POCI-01-0145-FEDER-030588, in part by the Regional OP Lisboa under Grant Lisboa-01-0145-FEDER-03058, in part by the FCT/MEC through National Funds under Projects MASSIVE5G under Grant PTDC/EEI-TEL/30588/2017, and in part by the OMIMO under Grant PTDC/EEI-TEL/30534/2017 and Grant UID/EEA/50008/2019.

ABSTRACT To fulfill 5th Generation (5G) communication capacity demands, the use of a large number of antennas has been widely investigated, and the array gain and spatial multiplexing that are offered by massive multiple input multiple output (mMIMO) have been used to improve the capacity. Fully digital architectures are not feasible for a large number of antennas, and hybrid analog/digital systems have emerged as options to retain a high number of antennas without as many radio frequency (RF) chains. However, these systems have, as consequences, non-avoidable nonlinear effects due to power amplifiers functioning in nonlinear regions. The strong nonlinear effects throughout the transmission chain will have a negative impact on the overall system's performance. Being able to access this impact is very important. For this purpose, we propose analytical and semi-analytical tools that allow for the evaluation of the nonlinear effects of a hybrid analog/digital orthogonal frequency-division multiplexing (OFDM) system. The proposed analysis starts with the characterization of the power amplifier's (PA) nonlinear response. This response is then used to derive a semi-analytic bit error rate expression. The theoretical tools are validated by using numerical results from two different cases: in the first one, the nonlinear PA response is assumed to follow an analytical model found in the literature and, in the second, the used nonlinear polynomial model mimics the response of a real amplifier. Using these two scenarios, the proposed tools are shown to be accurate making it possible to predict the nonlinearities' penalties in hybrid analog/digital OFDM systems and/or to assess the optimal operation point for a specific nonlinear amplifier.

INDEX TERMS Hybrid analogic/digital architectures, OFDM, massive MIMO, nonlinearities, power amplification.

I. INTRODUCTION

The explosive growth in wireless traffic and in the number of connected devices under user demands for high throughput and reliable communications, have resulted in the efficiency of current frequency bands being near their limits [1], [2]. A solution is to use a large number of antennas, which is referred to as massive multiple input multiple output (mMIMO), and has been identified as a key technology to handle orders of magnitude more data traffic [3], [4]. Superior capacities can be achieved with this technology through the rich and unique propagation diversity that is incorporated by the terminals, and adequate processing through precoding or beamforming. The mMIMO offers excellent spectral efficiency, which is achieved by the spatial multiplexing of many terminals in the same time/frequency resource. High-energy efficiency is also obtained in these systems through the array gain that permits a reduction of the radiated power [5], [6].

The mMIMO schemes cannot be regarded as scaled up versions of conventional MIMO schemes since the implementation and signal processing complexity would be prohibitively high [7]. In conventional MIMO systems, each

The associate editor coordinating the review of this manuscript and approving it for publication was Maurizio Magarini.

antenna has a fully dedicated radio frequency (RF) chain since all processing is completed at the digital level. However, fully digital architectures are not feasible for a large number of antennas due to hardware limitations, and thus hybrid analog/digital architectures have emerged as options to keep a high number of antennas without as many RF chains. In the recent years, hybrid analog-digital beamforming for both sub 6-GHz and millimeter wave bands were proposed [8]–[11], for which part of the signal processing is performed in the analog domain, and the reduced-complexity processing is left to the digital domain. Future cellular networks will likely use both sub 6-GHz and millimeter wave bands to accommodate different scenarios. In this paper, we mainly focus on the sub 6-GHz frequency bands. Previous works on hybrid beamforming for these frequencies can be found in [11]–[15]. In [11], a novel hybrid beamforming algorithm for mMIMO orthogonal frequency-division multiplexing (OFDM) systems with unified analog beamforming based on the spatial covariance matrix knowledge of all user terminals is proposed. A zero-forcing digital precoder and an unconstrained analog precoder that maximize the signal-to-leakage-plus-noise ratio are derived in [12]. Then, the authors developed a technique to design a constrained analog precoder that mimics the obtained unconstrained analog precoder under phase shifter constraints. In [13], several low-complexity transmission hybrid antenna selection and precoding schemes for narrowband massive MIMO systems were proposed. A new hybrid beamforming scheme for frequency-selective channels was designed in [14]. In this approach, the RF analog coefficient coherently combines the received signals from different time instants such that the energy of the desired symbol is focused onto a specific time sample. In [15], hybrid beamformers that minimize the estimation error in the data were designed for fully and partially connected architectures.

Beyond the high capacity demands, which can be fulfilled by using mMIMO technologies, the efficient utilization of limited bandwidth with high-data-rate transmissions while serving a large number of users is a prime requirement for present and future wireless communication systems. Aiming to meet this rising demand, OFDM systems are usually adopted due to their robustness in severely degraded channel conditions, link reliability and spectral efficiency. However, it is well known that the distortion in OFDM communications is severe due to the high peak-to-average power ratio (PAPR) of the input signal [16]. This causes some challenges to the power amplifier (PA) design since OFDM is very sensitive to nonlinear distortions [16]. PAs must stay in the linear region with a reasonable backoff in order to prevent strong nonlinear distortion effects. This means a reduced PA efficiency and smaller output power, which reinforces the importance of determining the distortion due to PAs [17].

Thus, research on calculating distortions and/or developing linearization techniques is crucial for these systems. Some works studied the nonlinearies (NLs) impact on conventional MIMO systems' performance [16], [18]–[20]. In [16], the importance of NL distortions due to transmitters' nonlinear PAs is highlighted, and its impact on a cooperative OFDM system is numerically evaluated. Linearization techniques have been proposed to mitigate distortion, though no theoretical analysis has been presented, and the work is relevant only for severely degraded channel conditions. In [18], a numerical study of PAs' impacts on single carrier MIMO communications with a suboptimal transmit beamforming scheme was performed. The performance of this system with different PA models was studied, although the study was limited to the numerical analysis of frequency-flat Rayleigh fading channels. The authors in [19] studied the ergodic achievable rate of a MIMO system under nonlinear PAs and showed that they can significantly degrade system performance. In [20], channel estimation and data transmitter link quality measurements were numerically simulated for OFDM communications, and three RF impairments were considered: oscillator phase, PA nonlinearity and in-phase/quadrature imbalance. The presented results were obtained only by simulation for a single user and a single antenna communication.

Recently, some works considered NL distortion effects in mMIMO systems. In the context of mMIMO with low-resolution ADCs, [21]–[23] presented an analysis of the distortion due to amplitude quantization, by using the Bussgang theorem. Concerning the effects of amplification in full digital mMIMO systems other works can be found [24]–[27]. In [24], authors used orthogonal Hermite polynomials to represent the amplified signal in mMIMO systems, which were partitioned into desired and distortion terms. The spatial crosscorrelation matrix of nonlinear distortion was derived, which could be applied to predict how the nonlinear distortion will behave in both in-band and out-band. In [25], the impact of nonlinear amplification was analytically derived. Accurate expressions for the spectral characterization of the transmission signals and bit error rate (BER) were derived. In [26], a polynomial model was considered to describe the nonlinear effects of the PAs and the Ito-Hermite polynomials were employed to derive the autocorrelation of the additional error term of the symbol estimate that is caused by the nonlinear PAs. In [27], a theoretical analysis of joint distortion, introduced by nonlinear low noise amplifiers, local oscillators with phase noise and oversampling finite resolution ADCs, was performed on the BS-side of uplink massive multiuser-MIMO OFDM systems.

The referred works did not consider hybrid analog/digital architectures and assumed one RF chain per antenna. However, fully digital architectures are difficult to implement in practical applications due to hardware limitations and the performance analysis of these works are not directly applicable for hybrid architectures. Therefore, the study of the NL effects in hybrid architectures is of paramount importance. The works [28]–[30] addressed the NL impact in hybrid analog/digital mMIMO architectures. In [28], the spectral and energy efficiencies of hybrid beamforming for narrowband mMIMO systems employing nonlinear PAs were numerically studied, and the covariance matrix for the distortion due to

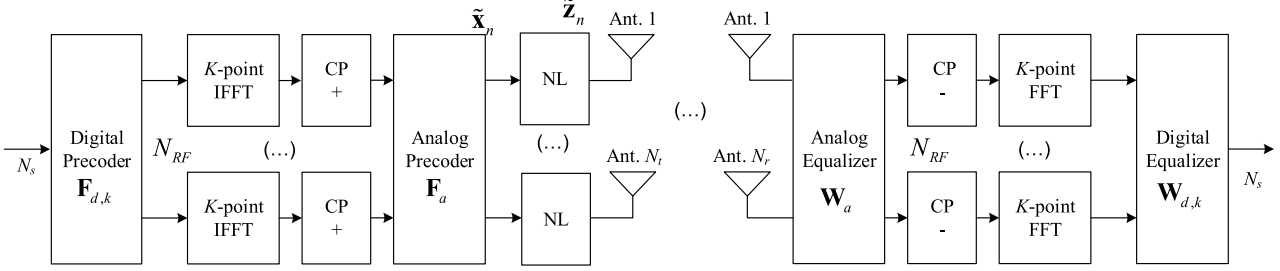

FIGURE 1. Transmitter and receiver block diagram.

the PAs is assessed. For the special case of a single RF chain, a closed form expression for the maximum spectral efficiency and for the lower bound on the achievable rate was presented. However, this work focused on the optimization of the energy efficiency of the system as a function of the consumed power per information bit, and did not consider the spectral characterization of the nonlinear distortion term. In [29], the spectral efficiency of a single carrier multiuser hybrid analog/digital architecture considering a PA modeled at the transmitter by a polynomial of order three was numerically evaluated. This was focused on the use of pre-distortion techniques to compensate nonlinear distortion effects, and does not present a theoretical analysis of the NL effects on the performance. In [30], a hybrid analog/digital OFDM mMIMO architecture is considered, and a theoretical analysis of the nonlinear distortion signal at the transmitter was conducted, although it did not assess the analytical impact in terms of BER performance.

Most of the previous works addressed the impact of NL in the context of either conventional MIMO or full digital mMIMO systems. The works that studied the impact of NL in hybrid analog/digital mMIMO are scarce and mostly assumed narrowband systems. A theoretical analysis of the impact in the performance of hybrid analog/digital mMIMO wideband systems has not yet been addressed, to the best of our knowledge. Therefore, in this paper, we aim to fill some of the gaps and propose analytical and semi-analytical tools that allow to efficiently evaluate the impact of PAs on a hybrid beamforming wideband mMIMO OFDM system. The impacts of nonlinearities, which are inherently present in these systems, are analytically and semi-analytically assessed with respect to the overall system performance. First, the distortion spectrum due to NLs is derived. Then, a semi-analytical expression for the BER using the previous derived distortion factor is derived, which, to the best of our knowledge, has not been addressed yet. This analysis reflects the propagation effect of the NL through the system chain. The proposed tools are valid for any nonlinear characteristic. To validate the derived tools, the impact of the PA nonlinearity in the hybrid mMIMO system was studied, considering two different PA modeling approaches. In the first case, an analytical model is used to characterize the NL characteristic of a PA. In the second one, a polynomial approximation is applied.

The remainder of the paper is organized as follows. Section II presents the system model of the hybrid analog/digital architecture considering OFDM and mMIMO technologies. The transmitter, receiver and the channel model are described in this section. In Section III, the PA models that are considered in the communication chain are described. The theoretical characterization of the NL distortion and BER expressions are derived in Section IV. Section V presents the main simulation results about the NL impacts, including the theoretical validation. The main conclusions are drawn in Section VI.

Notation

Throughout this paper, we will employ the following notation. Boldface capital letters denote matrices and boldface lowercase letters denote column vectors. The operation $(.)^H$ represents the Hermitian transpose of a matrix. The symbol $(*)$ represents the convolution operation. $\mathrm{Re}(c)$ and $\mathrm{Im}(c)$ represent the real and imaginary parts of c, respectively. $\mathbf{A}(j, l)$ denotes the element at row j and column $l$of matrix $\mathbf{A}$, and $\mathbf{a}(j)$ denotes element j of vector $\mathbf{a}$. $\mathbf{I}_N$ is the identity matrix with a size of $N \times N$. The symbols $\tilde{\mathbf{s}}$ and $\hat{\mathbf{s}}$ represent the time domain of the estimated signal relative to signal $\mathbf{s}$; $\mathbf{s}$ is the same variable in the frequency domain.

II. SYSTEM CHARACTERIZATION

The hybrid analog/digital mMIMO system that is considered in this work is presented in this section, and it is followed by the channel characterization.

A. SYSTEM MODEL

The hybrid analog/digital mMIMO OFDM system that is used in this work is shown in Fig. 1. We consider a single-user system with N_t transmission antennas and N_r receiving antennas, where the number of RF chains at both the transmitter and receiver is lower than the number of antennas, i.e., $N_{RF} \leq N_t$ and $N_{RF} \leq N_r$, respectively. An OFDM modulation with K available subcarriers is employed, where $K - K'$ are zero-values subcarriers, i.e., we consider an oversampling of factor$J = K/K'$.

The transmitted signal in the frequency domain before the NL is given by

$$\mathbf{x}_k = \mathbf{F}_a \mathbf{F}_{d,k} \mathbf{s}_k, \tag{1}$$

where $\mathbf{s}_k \in \mathbb{C}^{N_s \times 1}$ is the data vector and N_s is the number of data symbols that are transmitted in parallel over the N_t antennas for the kth subcarrier. We consider M-quadrature amplitude modulation (M-QAM) constellations in which the data symbols$\mathbf{s}_k$ have $\mathbb{E}\left[|s_k|^2\right] = \sigma_s^2 \mathbf{I}_{N_s}$. As it can be seen from (1), the transmitter processing is decomposed into two parts, the digital baseband and the analog circuitry, which are mathematically modeled using the precoder matrices $\mathbf{F}_a \in \mathbb{C}^{N_t \times N_{RF}}$ and $\mathbf{F}_{d,k} \in \mathbb{C}^{N_{RF} \times N_s}$, respectively. The digital part hasN_{RF} transmission chains and $N_s \leq N_{RF} \leq N_t$. First, the N_s data symbols are digitally precoded by using matrix $\mathbf{F}_{d,k}$ on each subcarrier. Then, a K-IFFT is employed on each RF chain followed by a cyclic prefix (CP). The time domain samples are further processed by using the analog coefficients, leading to the samples $\tilde{\mathbf{x}}_n$. Due to hardware constraints, the analog part is implemented using a matrix of analog phase shifters, which forces all elements of matrix $\mathbf{F}_a$ to have equal norms ($|\mathbf{F}_a(i_t, i_{rf})|^2 = N_t^{-1}$). The relation between the input signal of the NL block, $\tilde{\mathbf{x}}_n$, which is the inverse discrete Fourier transform (IDFT) of $\mathbf{x}_k$, and the output signal of the NL block, $\tilde{\mathbf{z}}_n$, will be discussed in Section III.

The received signal $\mathbf{y}_k \in \mathbb{C}^{N_r \times 1}$ at the subcarrier k is given by

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{z}_k + \mathbf{n}_k, \tag{2}$$

where $\mathbf{z}_k \in \mathbb{C}^{N_t \times 1}$ is the output signal of the NL in the frequency domain($\mathbf{z}_k$, which is the DFT of $\tilde{\mathbf{z}}_n$), which will be defined in the next section. $\mathbf{n}_k \in \mathbb{C}^{N_r \times 1}$ is zero mean Gaussian noise with a variance of σ_n^2 and $\mathbf{H}_k \in \mathbb{C}^{N_r \times N_t}$ is the channel matrix for subcarrier k. This received signal is first processed by using the analog phase shifters, which are modeled by the matrix $\mathbf{W}_a \in \mathbb{C}^{N_{RF} \times N_r}$.

All elements of matrix $\mathbf{W}_a$ must have equal norms. Then, baseband processing follows, which is composed of N_{RF} processing chains. It should be noted that the analog parts of both the precoder and equalizer are constant over all the subcarriers while the digital part is computed for each subcarrier. First, the CP is removed and a K-FFT is performed for each RF chain. Then, the signal passes through a linear digital filter $\mathbf{W}_{d,k} \in \mathbb{C}^{N_s \times N_{RF}}$. The filtered received signal is then given by $\mathbf{W}_{d,k}\mathbf{W}_a\mathbf{y}_k$, leading to the following estimates:

$$\hat{\mathbf{s}}_k = \mathbf{W}_{d,k}\mathbf{W}_a\mathbf{y}_k e^{-j\hat{\Theta}}, \tag{3}$$

where $\hat{\Theta}$ is the estimation of the phase term due to the amplitude modulation/phase modulation (AM/PM) function of PAs. The analog part of both the precoder and equalizer are constant over all the subcarriers, while the digital part is computed for each subcarrier. In this paper, we consider the hybrid analog-digital precoder and the hybrid equalizer in [30]. However, it should be emphasized that others hybrid precoder/equalizer recently proposed in the literature could also be considered, since our formulation is general and therefore it can be applied to any hybrid precoder/equalizer.

B. CHANNEL MODEL

The model that is used in the proposed system accurately models communications with tightly packed antenna arrays. The considered channel model follows the clustered frequency-selective channel model that is discussed in [9], which is represented as the sum of the contribution of N_{cl} clusters, each of which contributes with N_{ray} propagation paths to the channel matrix with a cyclic prefix duration, K_{CP}. The channel at subcarrier k is expressed as

$$\mathbf{H}_k = \sum_{d=0}^{K_{CP}-1} \mathbf{H}[d] e^{-k\frac{2\pi k}{K_c}d}, \tag{4}$$

where $\mathbf{H}[d]$ is the delay-d MIMO channel matrix with D taps, which can be written as

$$\mathbf{H}[d] = \lambda \sum_{i_c=1}^{N_{cl}} \sum_{i_r=1}^{N_{ray}} \alpha_{i_c,i_r} p\left(dT - \tau_{i_c,i_r}\right) \mathbf{A}_r(i_c, i_r)\mathbf{A}_t(i_c, i_r), \tag{5}$$

where α_{i_c,i_r} and τ_{i_c,i_r} correspond to the path gains and the relative time delay of the (i_r)th ray in the (i_c)th scattering cluster, respectively. λ is a normalization factor such that

$$\lambda = \sqrt{\frac{N_t N_r}{N_{ray} N_{cl}}}. \tag{6}$$

$p(\tau)$ is a pulse-shaping function for the T-spaced signaling that is evaluated at τ seconds [9]. $\mathbf{A}_t$ and $\mathbf{A}_r$ are the matrices of the array response vectors at the transmitter and receiver, which are formed by

$$\mathbf{A}_t = [\mathbf{a}_t(\theta_{t1,1}), \ldots, \mathbf{a}_t(\theta_{tN_{cl},N_{ray}}))], \tag{7}$$

and

$$\mathbf{A}_r = [\mathbf{a}_r(\theta_{r1,1}), \ldots, \mathbf{a}_r(\theta_{rN_{cl},N_{ray}}))], \tag{8}$$

respectively, where $\theta_{m,l}^r$ and $\theta_{m,l}^t$ are the azimuth angles of the arrival and departure, respectively. The array response vectors for an N_w-element uniform linear array are given by

$$\mathbf{a}_w(\theta_{wm,l}) = \frac{1}{\sqrt{N_w}}\left[1, e^{j\frac{2\pi d_a}{\lambda}\sin(\theta_{wm,l})}, \ldots, e^{j\frac{2\pi(N_v-1)d_a}{\lambda}\sin(\theta_{wm,l})}\right]^T, \tag{9}$$

where $w \in \{r, t\}$ and d_a is the inter-element spacing. Moreover, we assume that the duration of the delay-d MIMO channel is smaller than the cyclic prefix duration, K_{CP}, which is usually much lower than K. Therefore, we have $\mathbf{H}[d] = 0$ for $d > K_{CP}$.

III. NONLINEAR AMPLIFICATION MODELING

Although the presented analysis could easily be extended to other nonlinear sources, in this work, we assume that the nonlinearity presented in the overall transmitter is only due to the PA, which can cause severe linearity degradation. Due to efficiency reasons, PAs are normally operated near saturation,

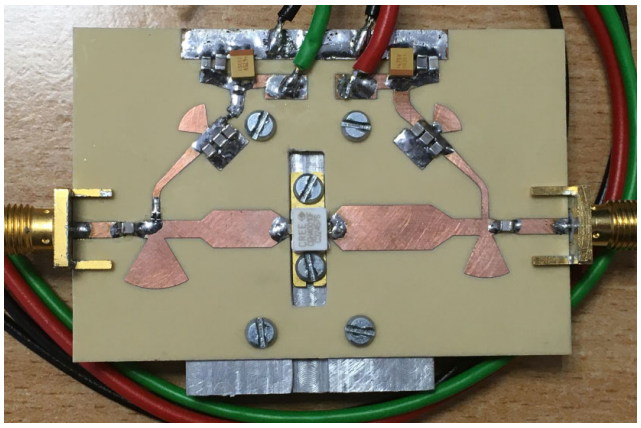

FIGURE 2. Photograph of the used 4.5GHz, 10W single-ended GaN PA.

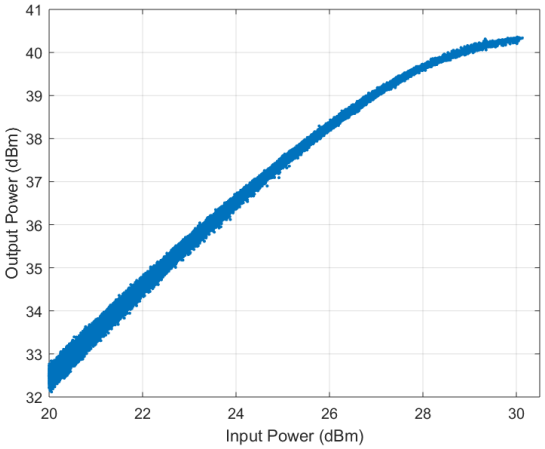

FIGURE 3. AM/AM characteristic of the used PA for a 10MHz OFDM input signal.

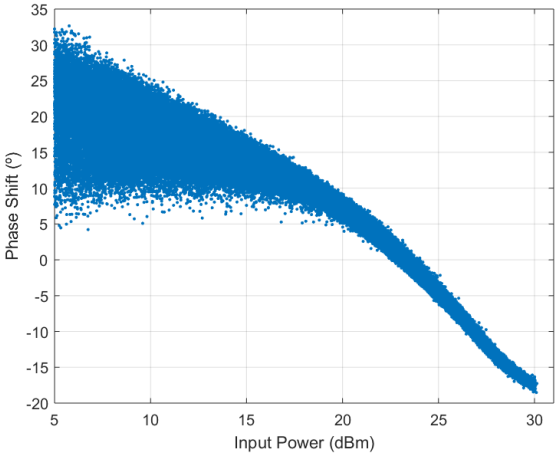

FIGURE 4. AM/PM characteristic of the used PA for a 10MHz OFDM input signal.

which increases signal distortion and justifies the effort spent on an accurate nonlinear characterization [31]. In this section we characterize the PA used, followed by the definition of the considered PA models.

A. PA CHARACTERIZATION

Power amplification at RF is most commonly achieved using travelling wave tube amplifiers (TWTA) or solid state power amplifiers (SSPA). SSPAs were mainly used for low- and medium-power levels due to size, operational lifetime and typically better performance, when compared to TWTAs. Modern communication systems use advanced PA architectures such as the Doherty or Chireix topologies [16].

In this work, to validate the proposed tool, we use a 4.5GHz 10W single-ended PA that incorporates a Wolfspeed's CGH40010F Gallium Nitrade (GaN) HEMT. Fig. 2 presents a photograph of that PA. The output matching network, built with microstrip lines, was designed so that the impedances presented to the transistor are the optimal ones at both the fundamental and the second harmonic (for the fundamental we choose the one that guarantees maximum efficiency and for the second harmonic a short circuit was applied). The input matching network, also built with microstrip lines, was designed to guarantee the necessary gain and assure stability at all frequencies [32].

Two of the most important estimators of the power amplification impact on communication systems linearity performance are the amplitude/amplitude (AM/AM) and the AM/PM conversion functions [33]. The nonlinear amplifier response transforms the time signal $\tilde{\mathbf{x}}_n$ into the signal $\tilde{\mathbf{z}}_n$, where $\tilde{\mathbf{x}}_n$ is the time-domain equivalent of the OFDM signal, i.e., $\left\{\tilde{\mathbf{x}}_n, n = 0, \ldots, K-1\right\} = \text{IDFT}\left\{\mathbf{x}_k, k = 0, \ldots, K-1\right\}$. The PA output can be given in function of the AM/AM and AM/PM functions, as follows

$$\tilde{\mathbf{z}}_n = f_A\left(r\right) e^{j\arg\left(\tilde{\mathbf{x}}_n\right) + jf_P(r)}, \tag{10}$$

where $r = \left|\tilde{\mathbf{x}}_n\right|, f_A\left(\cdot\right)$ and $f_P\left(\cdot\right)$ are the AM/AM and AM/PM nonlinear conversion functions, respectively. This signal thus includes the NL effect of the PA. The phase term, in (3), is an estimation of the phase rotation due to the AM/PM of the PAs, and is obtained through

$$\hat{\Theta} = \mathbb{E}\left[f_P\left(r\right)\right]. \tag{11}$$

Figures 3 and 4 show, respectively, the measured AM/AM and AM/PM characteristics of the used PA for a 10MHz OFDM input signal. With this signal, which has a peak to average power ratio (PAPR) of 9.1dB at the PA input and 7dB at the PA output, the amplifier provided an average drain efficiency of 39%.

B. POLYNOMIAL AND ANALYTICAL PA MODELS

The PA nonlinearity impact on a hybrid mMIMO system is derived in the next section, based in the assumption that the PA model is defined in terms of its AM/AM and AM/PM characteristics. Two different PA modeling approaches are used: polynomial and empirical analytical model. However, in principle, any other modeling approach could be used.

The use of polynomial models is nowadays quite disseminated throughout the behavioral modeling community since

they provide a low complexity, mathematically tractable and accurate characterization method [27], [28], [34]. The most common memory polynomial expression, expressed in its generic form, is given by

$$\tilde{\mathbf{z}}_n = \sum_{p=0}^{P-1} \sum_{m=0}^{M-1} a_{pm} \tilde{\mathbf{x}}_{n-m} \left|\tilde{\mathbf{x}}_{n-m}\right|^p \tag{12}$$

where a_{pm} are the model coefficients, $p+1$ is the polynomial order, P is the maximum polynomial order, m is the memory depth and M is maximum memory depth.

Note that when $M = 0$, the above model describes a purely static characteristic, defined as

$$\begin{aligned}\tilde{\mathbf{z}}_n &= \sum_{p=0}^{P-1} a_p \tilde{\mathbf{x}}_n \left|\tilde{\mathbf{x}}_n\right|^p = \sum_{p=0}^{P-1} a_p \left|\tilde{\mathbf{x}}_n\right|^{p+1} e^{j \arg(\tilde{\mathbf{x}}_n)} \\ &= f\left(\left|\tilde{\mathbf{x}}_n\right|\right) e^{j \arg(\tilde{\mathbf{x}}_n)},\end{aligned} \tag{13}$$

With $f\left(\left|\tilde{\mathbf{x}}_n\right|\right)$ being the global PA characteristic function. This signal can also be described as AM/AM and AM/PM functions, so that

$$f\left(\left|\tilde{\mathbf{x}}_n\right|\right) = f_A\left(\left|\tilde{\mathbf{x}}_n\right|\right) e^{jf_P(|\tilde{\mathbf{x}}_n|)}. \tag{14}$$

These polynomial models have been widely adopted because they are simpler to manipulate and are linear in the coefficients, meaning that these models can be easily fit by minimizing the mean squared error between model response and measurements. However, polynomial models may require many higher order terms. Instead, a generic baseband model can be used to represent these nonlinearities, being one of them the Rapp's model, which is simple but provides analysis tractability regardless of the used PA type [35], [36]. For this particular model $f_P(r)$ is approximately zero, and

$$f_A(r) = \frac{r}{\sqrt[2p]{\left(1 + \left(\frac{r}{s_M}\right)^{2p}\right)}}, \tag{15}$$

where p is the parameter that controls the smoothness of the transition between the linear and the saturation regions.

Figures 5 and 6 present, respectively the AM/AM and AM/PM obtained through measurements, polynomial approximation and empirical modeling approaches.

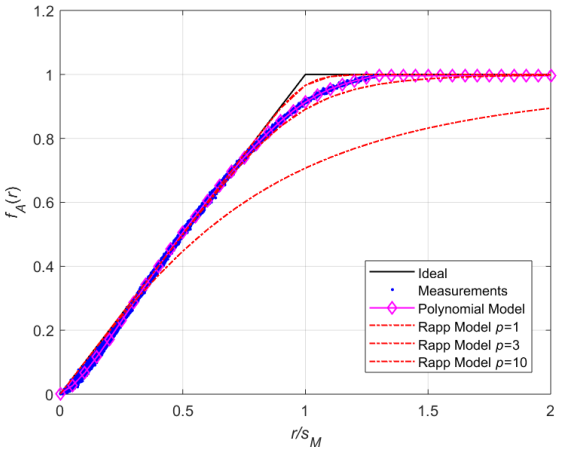

FIGURE 5. AM/AM conversion functions for the analytical and the polynomial PA models.

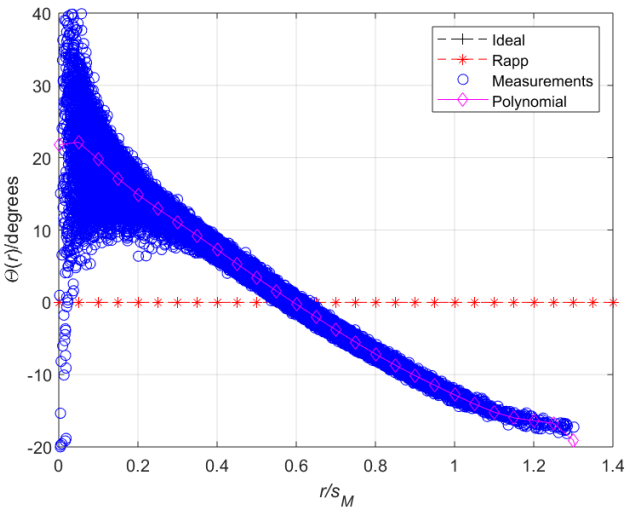

FIGURE 6. AM/PM conversion function of the analytical and polynomial PA models.

IV. THEORETICAL BER DERIVATION

To theoretically analyze the impact of the NLs on the overall hybrid mMIMO OFDM system that was previously defined, we start by analytically deriving the distortion at the output of the PA. Then, this distortion is used to conduct a semi-analytical BER analysis.

A. CHARACTERIZATION OF NL DISTORTION TERM

Concerning the distortion characterization, we start by defining the relation between the input and output signals of the PA. From central limit theorem, the input of the PA device is approximately Gaussian if the number of antennas and/or subcarriers is high. Therefore, from Bussgang's theorem, the output signal can be decomposed into uncorrelated useful and self-interference components [37], [38]. The component n of the PA output that is defined in (10) can be written as

$$\tilde{\mathbf{z}}_n = \alpha \tilde{\mathbf{x}}_n + \tilde{\mathbf{d}}_n \tag{16}$$

where $\tilde{\mathbf{d}}_n$ is the nonlinear distortion component, which is uncorrelated with $\tilde{\mathbf{x}}_n$, and

$$\alpha = \frac{1}{2\sigma^4} \int_0^{+\infty} r f_A(r) e^{jf_P(r)} e^{-\frac{r^2}{2\sigma^2}} dr \tag{17}$$

where $r = \left|\tilde{\mathbf{x}}_n\right|$. In addition, since the magnitude of $\tilde{\mathbf{x}}_n$ is a Rayleigh distribution, we can write

$$\alpha = \frac{\mathbb{E}\left[r f_A(r) e^{jf_P(r)}\right]}{\mathbb{E}\left[r^2\right]}. \tag{18}$$

The power spectral density (PSD) of the nonlinear input can be obtained by using the average values of the signal

that is associated with each subcarrier $G_{x,k} = \mathbb{E}\left[|x_k|^2\right]$. We assume that

$$G_{x,k} = \begin{cases} 2\sigma_x^2, & k = 0, \ldots, K' - 1 \\ 0, & k = K', \ldots, K - 1, \end{cases} \tag{19}$$

where $\mathbb{E}\left[|\mathbf{x}_k|^2\right] = \sigma_x^2 \mathbf{I}_{N_t}$, although the generalization to other cases is straightforward. For simplification purposes, $\tilde{x}_k$ is considered to be one element of the vector $\tilde{\mathbf{x}}_k$.

The autocorrelation of the input signal, which is defined as

$$R_{x,n} = \mathbb{E}\left[\tilde{x}_{n'} \tilde{x}^*_{n'-n}\right], n = 0, \ldots, K - 1, \tag{20}$$

where $n' \in \{0, 1, \ldots, K - 1\}$, can be obtained through the IDFT operator

$$\begin{aligned} &\left\{R_{x,n}, n = 0, 1, \ldots, K - 1\right\} \\ &\quad = \frac{K}{K'} IDFT\left\{G_{x,k}, k = 0, 1, \ldots, K' - 1\right\}, \end{aligned} \tag{21}$$

which, for component n, is given by the well-known expression

$$R_{x,n} = \frac{1}{K} \sum_{k=0}^{K} G_{x,k} e^{j2\pi k \frac{n}{K}}. \tag{22}$$

The autocorrelation of the nonlinearly distorted signal that is defined as $R_{nl,\Delta} = \mathbb{E}\left[\tilde{z}_n \tilde{z}^*_{n'}\right]$ can be expressed as a function of the input autocorrelation signal [30], as follows:

$$R_{NL,n} = 2 \sum_{\gamma=0}^{+\infty} I_{2\gamma+1} \frac{R_{x,n}^{\gamma+1} R_{x,n}^{*\gamma}}{R_{x,0}^{2\gamma+1}}, \tag{23}$$

where $I_{2\gamma+1}$ denotes the total power that is associated with the intermodulation product (IMP) of order $2\gamma + 1$, which is defined as

$$\begin{aligned} I_{2\gamma+1} = &\frac{1}{4\sigma^6 (\gamma + 1)} \\ &\times \left| \int_0^{+\infty} r^2 f_A(r) e^{j f_P(r)} e^{-\frac{r^2}{2\sigma^2}} L_\gamma\left(\frac{r^2}{2\sigma^2}\right) dr \right|^2, \end{aligned} \tag{24}$$

where the first intermodulation term corresponds to the useful power and where $L_\gamma^{(1)}(x)$ denotes the generalized Laguerre polynomial of order γ, which is defined as [39]

$$L_\gamma(x) = \sum_{k=0}^{n} (-1)^\gamma \frac{\gamma!}{(\gamma - k)! k! k!} x^k. \tag{25}$$

The PSD of the distorted signal due to the NL is given by

$$\begin{aligned} &\left\{G_{NL,k}, k = 0, \ldots, K' - 1\right\} \\ &\quad = \mathrm{DFT}\left\{R_{NL,n}, n = 0, \ldots, K - 1\right\}, \end{aligned} \tag{26}$$

where DFT represents the discrete Fourier transform operator. By replacing the $R_{NL,n}$ in (27) with (23) and using the property that an autocorrelation function has Hermitian symmetry, i.e., $\left(R_n^H\right)^* = R_{-n}^H$, the equivalent expression for the PSD of the NL signal can be simplified to

$$G_{NL,k} = 2 \sum_{\gamma=0}^{+\infty} \frac{I_{2\gamma+1}}{R_{NL,0}^{2\gamma+1}} g_{2\gamma+1}\left(G_{x,k}\right), \tag{27}$$

where

$$g_{2\gamma+1}(G_k) = \underbrace{(G_k * \ldots * G_k)}_{\gamma+1\text{times}} * \underbrace{(G_{-k} * \ldots * G_{-k})}_{\gamma\text{times}}, \tag{28}$$

since the exponent that denotes the $2\gamma + 1$ power is transformed into a $2\gamma + 1$ times convolution when applying the inverse Fourier transform. The convolutions are cyclic; however, if the oversampling is large enough, they become equivalent to linear convolutions.

The autocorrelation of the distortion signal that is defined as $R_{d,\Delta} = \mathbb{E}\left[\tilde{d}_n \tilde{d}^*_{n'}\right]$ can be expressed as a function of the input autocorrelation signal as follows:

$$R_{d,n} = 2 \sum_{\gamma=1}^{+\infty} I_{2\gamma+1} \frac{R_{x,n}^{\gamma+1} R_{x,n}^{*\gamma}}{R_{x,0}^{2\gamma+1}}. \tag{29}$$

The PSD of the distortion component can be obtained from

$$\begin{aligned} &\left\{G_{d,k}, k = 0, \ldots, K' - 1\right\} \\ &\quad = \mathrm{DFT}\left\{\left(\sum_{\gamma=1}^{+\infty} 2 I_{2\gamma+1} \frac{\left(R_{x,n}\right)^{\gamma+1} \left(R_{x,n}^*\right)^\gamma}{\left(R_{x,0}\right)^{2\gamma+1}}\right), \right. \\ &\qquad \left. \times n = 0, \ldots, K - 1 \right\}, \end{aligned} \tag{30}$$

and distortion power due to the NL is thus given by

$$\sigma_D^2 = \mathbb{E}\left[G_{D,k}\right], k = 1, \ldots, K'. \tag{31}$$

The diagram that is shown in Fig. 7 provides a simplified graphical view of the operations that are performed to obtain the statistical characterization of the NL distortion. By knowing the PSD of the PA input signal and the PA's AM/AM and AM/PM characteristic functions, it is possible to obtain the autocorrelation of the input signals and the IMP factors. Then, the calculation of the autocorrelation distortion, $R_{x,n}$, is obtained by using the autocorrelation of the input signal and the IMP coefficients. The DFT operation that is applied to the autocorrelation and is followed by the mean operator, being applied to the subcarriers, results in the distortion mean power.

B. BER ANALYSIS

In a second step of this theoretical analysis, the objective is to derive a semi-analytic expression for the BER. By replacing (1) and (2) in (3), the estimated signal of the kth subcarrier can be denoted as

$$\begin{aligned} \hat{\mathbf{s}}_k = &\alpha e^{-j\hat{\Theta}} \mathbf{W}_{d,k} \mathbf{W}_a \mathbf{H}_k \mathbf{F}_a \mathbf{F}_{d,k} \mathbf{s}_k + e^{-j\hat{\Theta}} \mathbf{W}_{d,k} \mathbf{W}_a \mathbf{H}_k \mathbf{d}_k \\ &+ e^{-j\hat{\Theta}} \mathbf{W}_{d,k} \mathbf{W}_a \mathbf{n}_k \end{aligned} \tag{32}$$

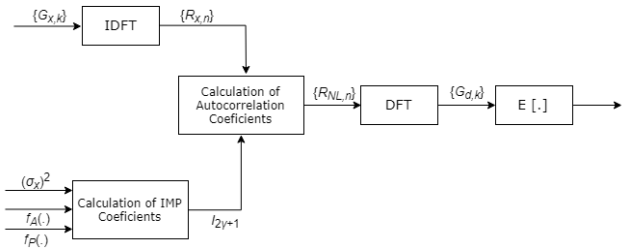

FIGURE 7. Operations that are performed to obtain the NL distortion.

By defining $\prod_k = \mathbf{W}_{d,k}\mathbf{W}_a\mathbf{H}_k\mathbf{F}_a\mathbf{F}_{d,k} \in \mathrm{C}^{N_s\times N_s}$, $\mathbf{A}_k = \mathbf{W}_{d,k}\mathbf{W}_a\mathbf{H}_k \in \mathrm{C}^{N_s\times N_t}$ and $\mathbf{W}_{eq,k} = \mathbf{W}_{d,k}\mathbf{W}_a \in \mathrm{C}^{N_s\times N_r}$, (33) can be simplified to

$$\hat{\mathbf{s}}_k = \alpha e^{-j\hat{\Theta}}\prod\nolimits_k \mathbf{s}_k + e^{-j\hat{\Theta}}\mathbf{A}_k\mathbf{d}_k + e^{-j\hat{\Theta}}\mathbf{W}_{eq,k}\mathbf{n}_k. \quad (33)$$

The expression for the ith estimated symbol becomes

$$\hat{\mathbf{s}}_k(i) = \alpha e^{-j\hat{\Theta}} \underbrace{\sum_{i_s=1}^{N_s}\prod\nolimits_k (i,i_s)\,\mathbf{s}_k(i_s)}_{\prod_k(i,i)\mathbf{s}_k(i)+\sum_{\substack{i_s=1\\ i_s\neq i}}^{N_s}\prod_k(i,i_s)\mathbf{s}_k(i_s)} + e^{-j\hat{\Theta}}\sum_{i_t=1}^{N_t}\mathbf{A}_k(i,i_t)\,\mathbf{d}_k(i_t) + e^{-j\hat{\Theta}}\sum_{i_r=1}^{N_r}\mathbf{W}_{eq,k}(i,i_r)\mathbf{n}_k(i_r), \quad (34)$$

with $i = 1,\ldots,N_s$.

From (34), we can see that the estimated symbol has three non-useful components: the interference between symbols, $\mathbf{e}_{int,k}$; the distortion due to NL, $\mathbf{e}_{dist,k}$; and the noise contribution, $\mathbf{e}_{noise,k}$. Therefore, the ith estimated symbol can be expressed as

$$\hat{s}_k(i) = \mathbf{u}_k(i) + \mathbf{e}_{int,k}(i) + \mathbf{e}_{dist,k}(i) + \mathbf{e}_{noise,k}(i), \quad (35)$$

with the useful component defined as

$$\mathbf{u}_k(i) = \alpha e^{-j\hat{\Theta}}\prod\nolimits_k (i,i)\,\mathbf{s}_k(i) \quad (36)$$

and each non-useful component defined, respectively, as

$$\mathbf{e}_{int,k}(i) = \sum_{\substack{i_s=1\\ i_s\neq i}}^{N_s}\alpha_k e^{-j\hat{\Theta}}\prod\nolimits_k (i,i_s)\,s_k(i_s), \quad (37)$$

$$\mathbf{e}_{dist,k} = \mathbf{A}_k(\mathbf{z}_k - \alpha\mathbf{x}_k) \quad (38)$$

and

$$\mathbf{e}_{noise,k} = e^{-j\hat{\Theta}}\mathbf{W}_{d,k}\mathbf{W}_a\mathbf{n}_k. \quad (39)$$

The BER is given by the Q-function of the ratio between the mean powers of the useful and non-useful components of the signal. To the best of our knowledge, a closed form expression for the calculation of the non-useful signal is not possible to obtain. It is well known that the noise contribution, $\mathbf{e}_{noise,k}$, follows a Gaussian distribution. Both interference and distortion variables, $\mathbf{e}_{int,k}$ and $\mathbf{e}_{dist,k}$, are also approximately Gaussian, since multicarrier signals with a large number of subcarriers are complex Gaussian distributed. Thus, an approach for the BER calculation is proposed for a given channel realization, where the non-useful signal power is calculated based on the power of the signals that are defined in (37), (38) and (39), as follows:

$$BER_k(i,\mathbf{H}_k)|_{\mathbf{H}} \simeq Q\left(\sqrt{\frac{\sigma_u^2}{\sigma_{int,k}^2+\sigma_{dist,k}^2+\sigma_{n,k}^2}}\right), \quad (40)$$

with the useful power given by

$$\sigma_{u,k}^2 = \alpha^2\sigma_s^2\left|\prod\nolimits_k (i,i)\right|^2. \quad (41)$$

The variances of the interference,$\sigma_{int,k}^2$; distortion, $\sigma_{dist,k}^2$; and noise, $\sigma_{nd,k}^2$ components are respectively given as functions of the variance of the signal s_k, σ_s^2; the distortion power, σ_D^2, as in (31); and the additive white Gaussian noise power that is added in the reception, σ_n^2. The three components of the non-useful power are approximated by the following expressions:

$$\sigma_{int,k}^2 \simeq \alpha_k^2\sigma_s^2\left(\left|\sum_{i'=1}^{N_s}\prod\nolimits_k (i,i')\right|^2 - \left|\prod\nolimits_k (i,i)\right|^2\right), \quad (42)$$

$$\sigma_{dist,k}^2 \simeq \sigma_D^2\sum_{i_t=1}^{N_t}|\mathbf{A}_k(i,i_t)|^2, \quad (43)$$

$$\sigma_{nd,k}^2 \simeq \sigma_n^2\sum_{i_r=1}^{N_r}\left|\mathbf{W}_{eq,k}(i,i_r)\right|^2. \quad (44)$$

Since $\prod_k$, $\mathbf{A}_k$ and $\mathbf{W}_{eq,k}$ are matrices that are dependent on the channel, the means in (42), (43) and (44) can be obtained through simulation to achieve channel realization.

The average BER for the kth subcarrier is then obtained by using the mathematical expectation of (40) for several channel realizations, including a total of N_{sim}different channels, which is given by

$$BER_k = \frac{1}{N_{sim}N_s}\sum_{j=1}^{N_{sim}}\sum_{i=1}^{N_s}BER_k\left(i,\mathbf{H}_{k,j}\right). \quad (45)$$

The block diagram in Fig. 8 summarizes the operations that are performed to achieve the BER for the presented hybrid analog/digital system. We start by calculating the distortion due to the PA (according to the steps in Fig. 7) and the factor α in (17). With these two parameters, in addition to the variance of the input PA signal and a matrix that is dependent on the channel state information (CSI), the useful signal power is calculated using (41). The signal power and a matrix based on the CSI is used to calculate the distortion component, while for the noise component the noise power and a matrix that is dependent on CSI are used.

The interference component is obtained by using the parameter α, the signal power and a matrix that is dependent

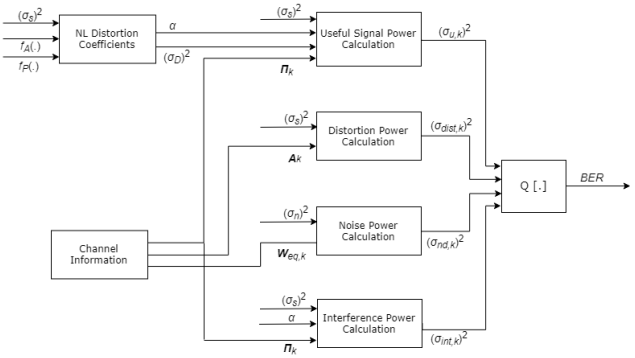

FIGURE 8. Operations that are performed to obtain the theoretical BER for a hybrid system.

on the CSI. Finally, these obtained values are used in the Q-function to obtain the semi-analytical BER expression, as seen in (40).

V. PERFORMANCE RESULTS

In this section, we present a set of performance results for the impact of NL on the hybrid digital/analog architecture that is described above. Our scenario has $N_{ray} = 10$ rays with Laplacian distributed azimuth angles of arrival and departure. The angular spread at both the transmitter and receiver is set to $\sigma = 10$ degrees. The antenna inter-element spacing d_a is assumed to be half-wavelength. The number of antennas at the transmitter and receiver are $N_r = 16$ and $N_t = 16$, respectively, and the number of RF chains is $N_{RF} = 8$. The number of subcarriers is set to $K = 256$ and the oversampling factor is set to 2. The number of data symbols that are transmitted per subcarrier is set to $N_s = 4$.

In this paper, we consider the hybrid analog-digital transmission and receiving beamforming that was proposed in [11]. However, it should be emphasized that other hybrid precoders/equalizers that have been recently proposed in the literature could also be considered since our formulation is general and could, therefore, be applied to any hybrid precoder/equalizer. We consider perfect oscillators, synchronization and channel estimation.

We assume that α_{i_c,i_l} are i.i.d. $\mathcal{CN}\left(0, \sigma^2_{\alpha,i_c}\right)$, where σ^2_{α,i_c} defines the average power of the i_cth cluster. The average power of all N_{cl} clusters is the same and $\sum_{i=1}^{N_{cl}} \sigma^2_{\alpha,i_c} = \lambda$, where λ is the normalization factor in (6).

The normalized spectra of signals associated to the PA can be seen in the following figures. In Fig. 9 the spectra of the normalized PA input and output signals are presented and in Fig. 10 the theoretical and simulated PA distortion signals are plotted, both for the case for which the oversampling factor is set to 2. In Fig. 11, we can see the spectra of the normalized PA input and output signals and in Fig. 12 the theoretical and simulated PA distortion signals, in these cases for an oversampling of 4. The presented spectra were obtained for different saturation levels $s_M/\sigma_x = 0.5$and $s_M/\sigma_x = 1$, for a set of channel realizations, so that we can observe the fluctuations in the signal due to the effects of the channel and

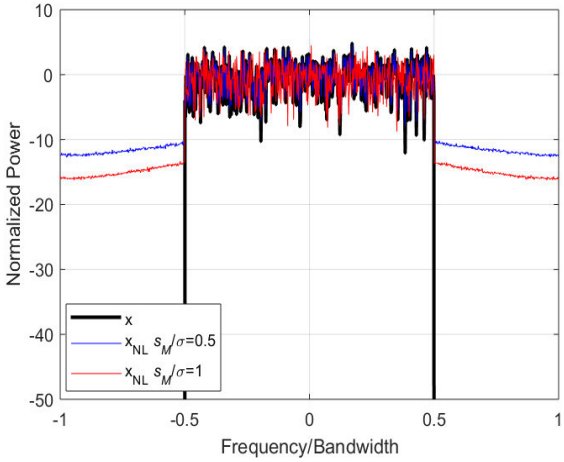

FIGURE 9. Normalized power spectrum of input and output signals of the PA when using the polynomial PA, with an oversampling factor of 2.

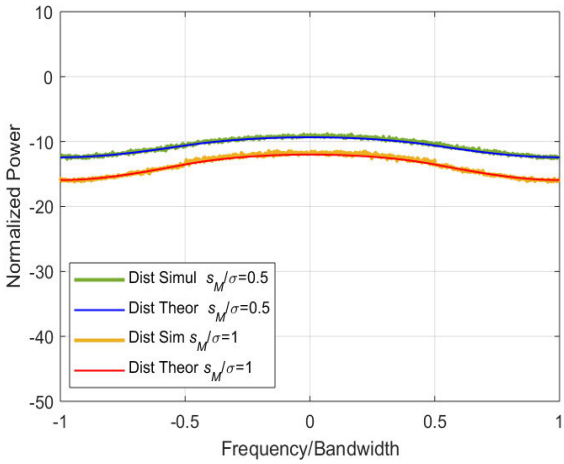

FIGURE 10. Normalized power of the distortion signals of the PA when using the polynomial PA, with an oversampling factor of 2.

the precoder. Thus, the signal distortion ratio is dependent on these signal fluctuations, since the distortion spectra are almost constant and the signals vary with the subcarriers. Moreover, the distortion spectrum variation along the frequency depends on the oversampling being almost constant for low oversampling values and decreasing in the non-useful band for the case of higher oversampling factors. Based in these results, we conclude that the oversampling should be higher than 2.

In the following figures, the impacts of the NL are shown in terms of the uncoded BER, which is presented as a function of E_b/N_0, where E_b denotes the average bit energy and N_0 denotes the one-sided noise power spectral density. An analysis in term of signal interference noise ratio (SINR) results, would lead to the same conclusions, accordingly to (40). The BER curves that are obtained theoretically, by using the proposed expressions, and numerically, by using Monte Carlo simulations, are plotted in Fig. 13 for the polynomial PA model, for saturation levels of $s_M/\sigma_x = 0.5$ and $s_M/\sigma_x = 1$. Moreover, the BER performances for similar conditions are presented in Fig. 14, in this case modeling the PA with Rapp's

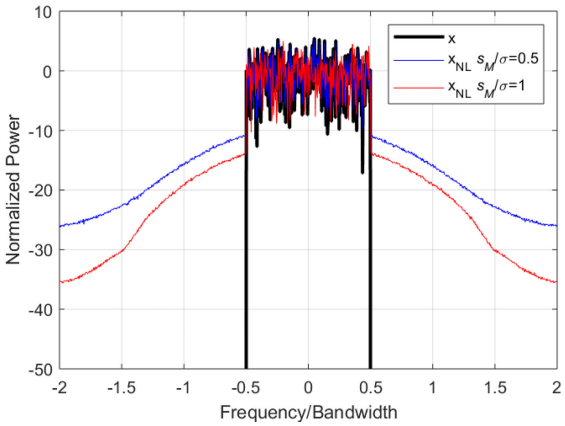

FIGURE 11. Normalized power of the input and output signals of the PA when using the polynomial PA with an oversampling factor of 4.

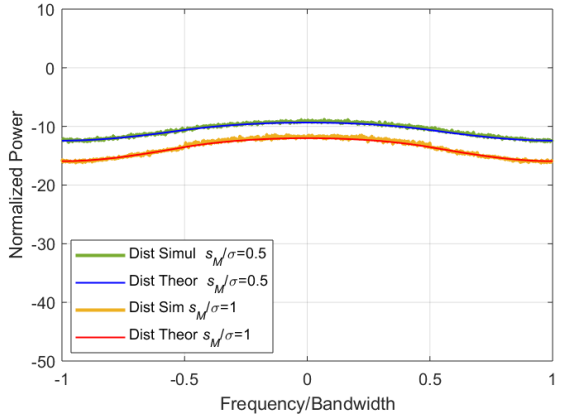

FIGURE 12. Normalized power of the distortion signals of the PA when using the polynomial PA with an oversampling factor of 4.

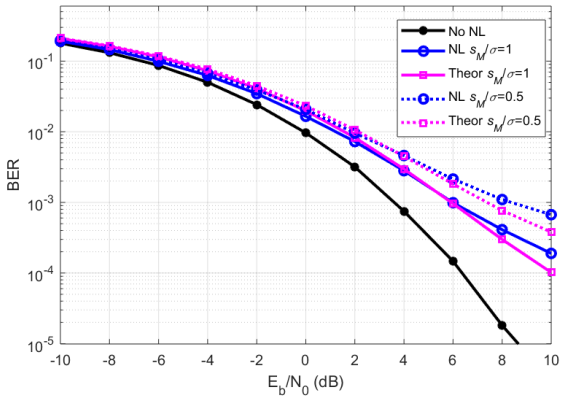

FIGURE 13. NL impacts on system in terms of the BER with different saturation levels and the polynomial PA model.

model and fixing the smoothness parameter as $p = 3$.The systems' performance is improved with the normalized saturation level. The closeness of the curves allows us to conclude about the derivation accuracy.

To observe the impacts of the PA, including the effects of working in the linear PA region, the total degradation value can be obtained. The total degradation variable, which fixed for a given BER, is defined as the sum of two components: one for the NL distortion and the other corresponding to the

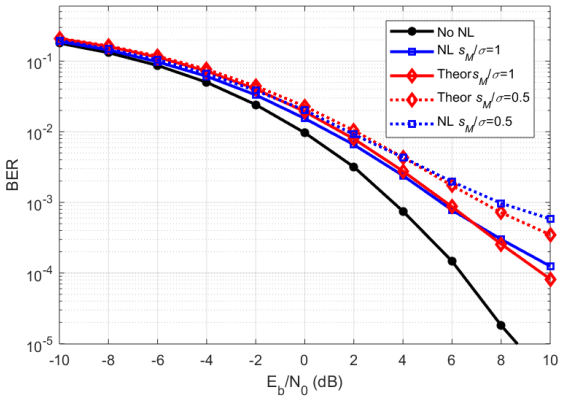

FIGURE 14. NL impacts on system in terms of the BER with different saturation levels, Rapp's model, and p =3.

effects of the amplifier functioning in the linear region. The corresponding expression is given by

$$D_{total}\left(s_M/\sigma\right) = D_{NL}\left(s_M/\sigma\right) + OBO, \tag{46}$$

where $D_{NL}\left(s_M/\sigma\right)$ is the degradation due to the NL, defined is for a fixed BER, and OBO is the value of the output backoff (OBO) power that corresponds to the power level at the output of the RF amplifier relative to the maximum output level possible using the RF amplifier, which is calculated as

$$OBO = \frac{P_{sat}}{P_{out}}. \tag{47}$$

Note that $P_{out} = \mathbb{E}\left[\left|\tilde{\mathbf{z}}_n\right|^2\right]$, which can easily be obtained by using

$$P_{out} = \int_0^\infty f_A^2(r)\, e^{jf_P(r)} p(r)\, dx, \tag{48}$$

where r is the absolute value of $\tilde{\mathbf{z}}_n$. The NL degradation is

$$D_{NL}\left(s_M/\sigma\right) = \gamma_{NL}\left(s_M/\sigma\right) - \gamma_L, \tag{49}$$

where γ_{NL} is the E_b/N_0 that corresponds to a fixed BER and γ_L is the observed ratio E_b/N_0 for the same communication for a target BER but without the PA.

Fig. 15 presents the total degradation and the two components that form it for the polynomial PA model and for a BER target $BER = 10^{-3}$. One of the components is the OBO power, which is defined as the ratio between the saturation level power and the PA output power. OBO power increases with saturation level, as observed. The other component is the degradation in E_b/N_0 due to the NL, which decreases as the saturation level increases, as was already observed in Fig. 13.

The optimal operating point is the one with the lowest degradation in E_b/N_0, which for BER=10^{-3}, is approximately $s_M/\sigma_x = 0.7$. This is the case that best compensates for the power degradation that is caused by being in the linear region and the distortion that is generated by the PA (that decreases as the saturation value increases). As the target BER decreases, the saturation level that produces lower distortion increases, since the BER does not significantly differ

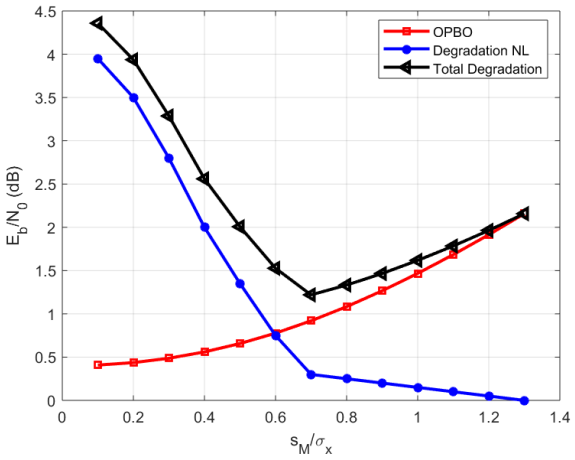

FIGURE 15. Total degradation as a function of the saturation level for $BER = 10^{-3}$ and polynomial PA model.

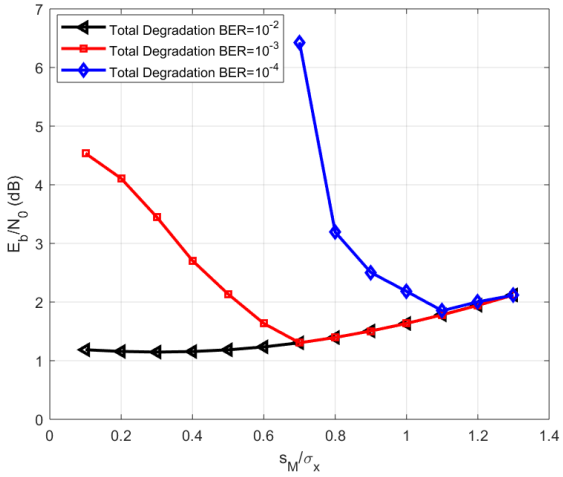

FIGURE 16. Total degradation as a function of the saturation level for different BERs and polynomial PA model.

with or without the NL for high E_b/N_0 values. For other BER targets, the optimal saturation levels can be obtained from Fig. 16. For BER=10^{-4}, the optimal point is $s_M/\sigma_x = 1.1$, and for BER=10^{-2} it is $s_M/\sigma_x = 0.3$.

VI. CONCLUSION

In this paper, a hybrid analog/digital mMIMO OFDM system is studied. To analyze the effects of nonlinearities on the performance of massive MIMO systems, we derived theoretical tools to estimate the distortion at the output of the NL and the BER at the detection level.

The proposed theoretical and semi-analytical tools were successfully validated through Monte Carlo simulations, in which the NLs were modeled using analytical and polynomial PAs' models. We concluded that the nonlinearity in the transmitter, which was due to hardware impairments, had a strong impact on hybrid analog/digital mMIMO systems, as expected. Thus, the proposed distortion derivation has significant application prospects for hybrid broadband wireless systems.

We also showed that the useful signal spectrum fluctuates differently for different subcarriers; thus, the lower power subcarriers will be strongly affected by the distortions. Moreover, it is possible to obtain the optimal operating point for a specific nonlinear amplifier, which can be useful in communication systems' development.

The knowledge of the error due to nonlinear amplifiers during detection is of huge importance since it can be used in the precoding and/or equalizer design in order to reduce the observed penalty. It would be interesting to extend the proposed tool for millimeter wave systems. However, this requires appropriate PA models, and thus it will be object of study in further work.

REFERENCES

[1] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO has unlimited capacity," *IEEE J. Wireless Commun.*, vol. 17, no. 1, pp. 574–590, Jan. 2017.

[2] M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Zhu, P. De Silva, F. Tufvesson, A. Benjebbour, and G. Wunder, "5G: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1201–1221, Jun. 2017.

[3] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: Ten myths and one critical question," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 114–123, Feb. 2016.

[4] K. Zheng, L. Zhao, J. Mei, B. Shao, W. Xiang, and L. Hanzo, "Survey of large-scale MIMO systems," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1738–1760, 3rd Quart., 2015.

[5] E. G. Larsson and L. Van Der Perre, "Massive MIMO for 5G," *IEEE 5G Tech Focus*, vol. 1, no. 1, pp. 1–4, Mar. 2017.

[6] T. E. Bogale and L. B. Le, "Massive MIMO and mmWave for 5G wireless HetNet: Potential benefits and challenges," *IEEE Veh. Technol. Mag.*, vol. 11, no. 1, pp. 64–75, Mar. 2016.

[7] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[8] S. Han, C.-L. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.

[9] A. Alkhateeb and R. W. Heath, Jr., "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, May 2016.

[10] R. Magueta, D. Castanheira, A. Silva, R. Dinis, and A. Gameiro, "Hybrid iterative space-time equalization for multi-user mmw massive MIMO systems," *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 608–620, Feb. 2017.

[11] D. Zhu, B. Li, and P. Liang, "A novel hybrid beamforming algorithm with unified analog beamforming by subspace construction based on partial CSI for massive MIMO-OFDM systems," *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 594–607, Feb. 2017.

[12] S. Park, J. Park, A. Yazdan, and R. W. Heath, Jr., "Exploiting spatial channel covariance for hybrid precoding in massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 14, pp. 3818–3832, Jul. 2017.

[13] P. V. Amadori and C. Masouros, "Large scale antenna selection and precoding for interference exploitation," *IEEE Trans. Commun.*, vol. 65, no. 10, pp. 4529–4542, Oct. 2017.

[14] S. Payami, M. Sellathurai, and K. Nikitopoulus, "Low-complexity hybrid beamforming for massive MIMO systems in frequency-selective channels," *IEEE Access*, vol. 7, pp. 36195–36206, 2019.

[15] S. S. Ioushua and Y. C. Eldar, "A family of hybrid analog–digital beamforming methods for massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 12, pp. 3243–3257, Jun. 2019.

[16] P. K. Singya, N. Kumar, and V. Bhatia, "Mitigating NLD for wireless networks: Effect of nonlinear power amplifiers on future wireless communication networks," *IEEE Microw. Mag.*, vol. 18, no. 5, pp. 73–90, Jul./Aug. 2017.

[17] F. H. Raab, P. Asbeck, S. Cripps, P. B. Kenington, Z. B. Popovic, N. Pothecary, J. F. Sevic, and N. O. Sokal, "Power amplifiers and transmitters for RF and microwave," *IEEE Trans. Microw. Theory Techn.*, vol. 50, no. 3, pp. 814–826, Mar. 2002.

[18] J. Qi and S. Aïssa, "On the power amplifier nonlinearity in MIMO transmit beamforming systems," *IEEE Trans. Commun.*, vol. 60, no. 3, pp. 876–887, Mar. 2012.
[19] M. Fozooni, M. Matthaiou, E. Björson, and T. Q. Duong, "Performance limits of MIMO systems with nonlinear power amplifiers," in *Proc. GLOBECOM*, San Diego, CA, USA, Dec. 2015, pp. 1–7.
[20] Y. Zou, P. Zetterberg, U. Gustavsson, T. Svensson, A. Zaidi, T. Kadur, W. Rave, and G. Fettweis, "Impact of major RF impairments on mm-Wave communications using OFDM waveforms," in *Proc. IEEE Globecom*, Washington, DC, USA, Dec. 2016, pp. 1–7.
[21] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "Throughput analysis of massive MIMO uplink with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 4038–4051, Jun. 2017.
[22] C. Mollén, J. Choi, E. G. Larsson, and R. W. Heath, Jr., "Uplink performance of wideband massive MIMO with one-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 87–100, Oct. 2016.
[23] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. Swindlehurst, and L. Liu, "Channel estimation and performance analysis of one-bit massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4075–4089, Aug. 2017.
[24] C. Mollén, U. Gustavsson, T. Eriksson, and E. G. Larsson, "Spatial characteristics of distortion radiated from antenna arrays with transceiver nonlinearities," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6663–6679, Oct. 2018.
[25] J. Guerreiro, R. Dinis, and P. Montezuma, "Massive MIMO with nonlinear amplification: Signal characterization and performance evaluation," in *Proc. Globecom*, Washington, DC, USA, Dec. 2016, pp. 1–6.
[26] C. Mollén, U. Gustavsson, T. Eriksson, and E. G. Larsson, "Impact of spatial filtering on distortion from low-noise amplifiers in massive MIMO base stations," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6050–6067, Dec. 2018.
[27] S. Jacobsson, U. Gustavsson, G. Durisi, and C. Studer, "Massive MU-MIMO-OFDM uplink with hardware impairments: Modeling and analysis," in *Proc. 52nd Asilomar Conf. Signals, Syst., Comput.*, Oct. 2018, pp. 1829–1835.
[28] N. N. Moghadam, G. Fodor, M. Bengtsson, and D. J. Love, "On the energy efficiency of MIMO hybrid beamforming for millimeter-wave systems with nonlinear power amplifiers," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 7208–7221, Nov. 2018.
[29] H. Yan and D. Cabric, "Digital predistortion for hybrid precoding architecture in millimeter-wave massive MIMO systems," in *Proc. Int. Conf. Acoust., Speech Signal Process.*, New Orleans, LA, USA, Mar. 2017, pp. 3479–3483.
[30] S. Teodoro, A. Silva, R. Dinis, and A. Gameiro, "Performance impact of nonlinear amplification in massive MIMO mmWave systems," in *Proc. IEEE VTC-Spring*, Kuala Lumpur, Malaysia, Apr./May 2019, pp. 1–5.
[31] P. M. Lavrador, T. R. Cunha, P. M. Cabral, and J. C. Pedro, "The linearity-efficiency compromise," *IEEE Microw. Mag.*, vol. 11, no. 5, pp. 44–58, Aug. 2010.
[32] S. C. Cripps, *Advanced Techniques in RF Power Amplifier Design*. Norwood, MA, USA: Artech House, 2002.
[33] L. Cotimos Nunes, P. M. Cabral, and J. C. Pedro, "AM/AM and AM/PM distortion generation mechanisms in Si LDMOS and GaN HEMT based RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 4, pp. 799–809, Apr. 2014.
[34] 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.
[35] J. Joung, C. K. Ho, K. Adachi, and S. Sun, "A survey on power-amplifier-centric techniques for spectrum- and energy-efficient wireless communications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 315–333, 1st Quart., 2014.
[36] C. Rapp, "Effects of HPA-nonlinearity on a 4-DPSK/OFDM-signal for a digital sound broadcasting signal," in *Proc. 2nd Eur. Conf. Satell. Commun.*, Liege, Belgium, Oct. 1991, pp. 179–184.
[37] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted Gaussian signals," Res. Lab. Elec., Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep. 216, Mar. 1952.
[38] R. Price, "A useful theorem for nonlinear devices having Gaussian inputs," *IRE Trans. Inf. Theory*, vol. 4, no. 2, pp. 69–72, Jun. 1958.
[39] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*. New York, NY, USA: Dover, 1972.

SARA TEODORO received the Licenciatura degree in electrical and computer engineering from the University of Coimbra, in 2004, and the Ph.D. degree from the Department of Electronics and Telecommunications, University of Aveiro, in 2011. She has an Assistant Lecturer experience with the Department of Electronics, Polytechnic Institute of Leiria, from 2005 to 2008. She is currently a Postdoctoral Researcher with the Instituto de Telecomunicações, Pólo de Aveiro. Her research work was in cooperative diversity and distributed space-time/frequency codes for OFDM systems, in interference alignment, and quantization techniques for wireless communications. Her current interests are in linearization techniques applied to hybrid analog/digital architectures, for massive MIMO and millimeter wave systems.

ADÃO SILVA received the M.Sc. and Ph.D. degrees in electronics and telecommunications from the University of Aveiro, in 2002 and 2007, respectively. He is currently an Assistant Professor with the Department of Electronics, Telecommunications and Informatics, University of Aveiro, and a Senior Researcher with the Instituto de Telecomunicações. He has been participating in several national and European projects, namely, the ASILUM, MATRICE, 4MORE within the ICT programme, and the FUTON and CODIV projects with the FP7 ICT. He has led several research projects, in the broadband wireless communications area, at the national level. His interests include multiuser MIMO, multicarrier-based systems, cooperative networks, precoding, multiuser detection, massive MIMO, and millimeter wave communications. He acted as a member of the TPC of several international conferences.

RUI DINIS received the Ph.D. degree from the Instituto Superior Técnico (IST), Technical University of Lisbon, Portugal, in 2001, and the Habilitation degree in telecommunications from the Faculdade de Ciências e Tecnologia (FCT), Universidade Nova de Lisboa (UNL), in 2010. He was a Researcher with the Centro de Análise e Processamento de Sinal (CAPS), IST, from 1992 to 2005. From 2001 to 2008, he was a Professor at IST. In 2003, he was an Invited Professor with Carleton University, Ottawa, ON, Canada. He was also a Researcher with the Instituto de Sistemas e Robótica (ISR), from 2005 to 2008. Since 2009, he has been a Researcher with the Instituto de Telecomunicações (IT). He is currently an Associate Professor with FCT-UNL. He has been actively involved in several national and international research projects in the broadband wireless communications area. His research interests include modulation, equalization, channel estimation, and synchronization. He was also a Guest Editor for *Physical Communication* (Elsevier) (Special Issue on Broadband Single-Carrier Transmission Techniques). He is also an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS (Transmission Systems-Frequency-Domain Processing and Equalization) and the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.

FILIPE M. BARRADAS (S'13–M'17) was born in Évora, Portugal, in July 1989. He received the M.Sc. degree in electronics and telecommunications engineering and the Ph.D. degree in electrical engineering from the Universidade de Aveiro, Aveiro, Portugal, in 2012 and 2017, respectively. He is currently a Research Assistant with the Instituto de Telecomunicações, Aveiro, Portugal. His main interests include digital predistortion and behavioral modeling of RF PAs, as well as signal processing with applications on telecommunications, and design and analysis of nonlinear microwave circuits. He has been a Reviewer of several IEEE journals.

PEDRO M. CABRAL (S'04–M'07–SM'16) was born in Aveiro, Portugal, in 1979. He received the M.Sc. and Ph.D. degrees from the Universidade de Aveiro, Aveiro, Portugal, in 2002 and 2006, respectively. He is currently a Senior Researcher with the Instituto de Telecomunicações, Aveiro, and an Assistant Professor with the Universidade de Aveiro. His current research interests include active device nonlinear modeling, design of microwave circuits, high-efficiency power amplifiers, and wireless transmitter architectures. He has been a Reviewer of several IEEE publications, including the IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES, the IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS, the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—I: REGULAR PAPERS.

ATÍLIO GAMEIRO received the Licenciatura and Ph.D. degrees from the University of Aveiro, in 1985 and 1993, respectively. He is currently an Associate Professor with the Department of Electronics and Telecommunications, University of Aveiro, and a Researcher with the Instituto de Telecomunicações, Pólo de Aveiro, where he is also the Head of the group. His industrial experience includes a period of one year with BT Labs and one year with NKT Elektronik. His current research interests include signal processing techniques for digital communications and communication protocols, and within this research line he has done work for optical and mobile communications, either at the theoretical and experimental level. He has published more than 200 technical articles in international journals and conferences. His current research activities involve space-time-frequency algorithms for the broadband wireless systems and cross-layer design. He has been involved and has led IT or the University of Aveiro participation on more than 20 national and European projects.