



**HAL**  
open science

# End-to-end deep learning for multipair two-way massive MIMO with PA impairments

Maha Cherif, Ahlem Arfaoui, Rafik Zayani, Ridha Bouallegue

## ► To cite this version:

Maha Cherif, Ahlem Arfaoui, Rafik Zayani, Ridha Bouallegue. End-to-end deep learning for multipair two-way massive MIMO with PA impairments. *IEEE Systems Journal*, 2023, 17 (2), pp.3150-3159. 10.1109/JSYST.2022.3232590 . cea-04301939

**HAL Id: cea-04301939**

**<https://cea.hal.science/cea-04301939>**

Submitted on 23 Nov 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# End-to-End Deep Learning for Multipair Two-Way Massive MIMO with PA Impairments

Maha Cherif , Ahlem Arfaoui, Rafik Zayani , and Ridha Bouallegue

**Abstract**—Multipair (MP) massive multiple-input–multiple-output (MIMO) two-way relaying system is a promising solution that is able to provide excellent spectral efficiency performance. It consists in multiple pairs ( $2K$ ) of single-antenna user terminals that exchange information through of a relay using a large number of antennas  $M_t \gg 2K$ . The relay is equipped with a large number of energy-efficient power amplifiers (PAs) and ensure the multipair two-way forwarding. In this article, we first present an analytical study focused on the effects of PAs on the considered system over block fading Rayleigh channel. We derive closed-form expressions for the symbol error rate. Then, we develop a new efficient distributed learning approach for MP two-way massive MIMO. Specifically, we design the MP two-way massive MIMO system by leveraging the concept of the autoencoder, whereas the end-to-end communication system is designed using one deep neural network (NN). Interestingly, the proposed scheme, referred to as AE-MP-mMIMO, is suited for varying block fading Rayleigh channel scenarios. Indeed, we propose to adopt one stage precoder/decoder for each UE and two-stage precoding scheme for the relay: 1) an NN-Tx is used to address the PA nonlinearity and 2) a linear zero-forcing precoder is adopted to remove the multiuser interference. The NN-Tx and the UE's precoder/decoder are trained off-line and when the channel varies, only the zero-forcing precoder changes on-line. Numerical simulations show the capability of the proposed approach to achieve competitive performance with a significantly lower complexity compared to existing literature.

**Index Terms**—Multipair, massive multiple-input–multiple-output (MIMO), power amplifier, deep learning (DL), end to end (E2E).

## I. INTRODUCTION

MASSIVE multiple-input–multiple-output (MIMO) system is a promising technology for the sixth generation (6G) wireless communication standard. Indeed, massive MIMO represents a multiuser (MU)-MIMO technology where each base station (BS) is equipped with a large number of transmit antennas, providing a service to a large number of users [1]. This new system is regarded as an efficient and scalable approach to improve both the spectral-efficiency (SE) and energy-efficiency (EE). Several studies presented in [2] and [3] have shown that massive MIMO can provide good performance in time division

duplex (TDD) mode where the uplink and downlink transmissions take place in the same frequency resource but are separated in time.

Massive MIMO systems typically adopt direct linear signal processing methods, such as the matched filter (MF) and zero forcing (ZF), because of their ability to reduce transmitter power consumption, to realize signal transmission and to eliminate multiuser interference [4]. However, these linear precoders carry high peak-to-average power ratio (PAPR) to transmit signals [5]. Consequently, massive MIMO signals are very sensitive to the nonlinearity of the radio frequency (RF) power amplifier (PA), which is the main hardware distortion and is inevitable in a typical transmission chain. Indeed, when RF PA is present in the transmission chain, signal distortion and phase rotation emerge, resulting in a transmission performance degradation.

PAs give enough output power at the transmitter level, especially at the BS, to ensure the routing of the communication over a long distance. Except that, they are the main source of hardware impairments (HWIs) and are expected to be low-cost and energy-efficient component to enable cost and energy efficient massive MIMO BS deployments.

## A. Related Works

Many works have aimed at interpreting the two-way relaying technology, aiming to allow some pairs of users to communicate through an intermediate relay. At this level, the relay plays a crucial role in solving the resource allocation problem between an integrated device-to-device (D2D) terminal pair in a time division duplex cellular network, which was adopted by the IEEE 802.11ax [6] to maximize spectrum efficiency.

Recently, a lot of academic and industrial researchers have focused on studying the SE of multipair unidirectional [7] and bidirectional two-way half-duplex massive MIMO relaying [8], for different power scaling schemes. Authors in [9] derived the asymptotic expressions of the SE and EE of the MP system when the number of relay antennas goes to infinity. In [10], authors studied the joint effect of hardware impairments and imperfect channel state information (CSI) on multirelay networks.

Note that no explicit analytical performance evaluation of MP two-way massive MIMO system under PA nonlinearity has never been presented in the open literature, motivating us to establish an analytical study in terms of symbol error rate (SER), giving insight into the system deployment and its power efficiency. In order to improve the PA linearity and efficiency, two complementary solutions have mainly been studied in literature which are PA linearization [11] and PAPR reduction [12].

On one hand, digital predistortion (DPD) has been shown as the most promising solution for PA linearization, providing excellent performance. It consists in adding a module before the

Manuscript received 1 April 2022; revised 19 September 2022; accepted 22 December 2022. (Corresponding author: Maha Cherif.)

Maha Cherif and Ahlem Arfaoui are with the Innov'COM Lab., SupQ2 Com, Ariana 2083, Tunisia (e-mail: maha.cherif@gmail.com; ahlemarfaoui07@gmail.com).

Rafik Zayani is with the CEA-Leti, Université Grenoble Alpes, 38054 Grenoble, France (e-mail: rafik.zayani@cea.fr).

Ridha Bouallegue is with the SupCom, El Ghazala, Ariana 2083, Tunisia (e-mail: ridha.bouallegue@gmail.com).

Digital Object Identifier 10.1109/JSYST.2022.3232590

power amplifier, such that the resulting system, i.e., DPD plus PA, is linear. Nevertheless, this solution can not be adequate to massive MIMO systems because of the extremely high number of DPD modules needed in such massive MIMO BS.

On the other hand, the most important PAPR reduction techniques are tone reservation (TR) [13], selective mapping (SLM) [14], iterative clipping and filtering [15], active constellation extension [16], and coding [17]. These methods can, unfortunately, provide moderate PAPR reduction that does not satisfy the requirements in high power-efficiency for massive MIMO systems.

Recently, more interesting techniques have been proposed for massive MIMO systems, like joint MU precoding and PAPR reduction schemes introduced in [5] and [18], referred to as MU-PPGDm, RZF-OPNS, POLY-POLY-Horner, and POLY-OPNS. They consist in gradient-iterative method-based linear precoding. Unfortunately, all of these techniques were limited to traditional/colocated massive MIMO systems. Although good performance is offered by these schemes, they suffer from some drawbacks like increase of the average power (i.e., signal-to-noise ratio gap), high latency, and high computational complexity.

Regarding these observations, it is of paramount importance to study a cutting edge approach that can enable excellent energy-efficient massive MIMO system while requiring low computational complexity.

More recently, deep learning (DL) has given significant success in a wide set of applications. Interestingly, it has been shown to be able to provide efficient solutions for physical layer of wireless communications. In [19], a relatively comprehensive study of DL techniques for wireless communication systems was interpreted. It is worth citing channel estimation and signal detection based DL studied in [20] and [21]. Most importantly, end-to-end learning of communications systems has attracted a lot of interest within the wireless community. Its goal is to learn full communication system, i.e., transmitter, channel, and receiver, which are optimized for specific performance metrics and a channel model. A simple framework proposed in [22] consists in interpreting the whole communication system as an *autoencoder*. This latter has been also adopted for MIMO communications where we can cite the work in [23] that introduced a novel end-to-end learning approach for uplink multiuser single-input–multiple-output (SIMO) joint transmitter and noncoherent receiver design in fading channels.

## B. Motivation

Different observations have motivated this work, we can cite: 1) The MP two-way massive MIMO is an emerging technology providing orders of magnitude gains in terms of spectral-efficiency and energy-efficiency, targeting the 6G goals. 2) MP two-way massive MIMO can be economically attractive only if it adopts low-cost hardware, leading to severe HWI which is very challenging to mitigate. 3) DL together with the latest advancements of massive MIMO's precoding schemes can offer radical performance improvements, leading to practical and energy-efficient MP two-way massive MIMO. 4) Green signal processing solutions become extremely needed w.r.t. the sustainability of the telecommunication sector and decreasing carbon footprint.

## C. Contributions

Regarding the issues discussed above and this work's motivation, behind this work, our main contributions are summarized as follows.

- 1) We established an original theoretical performance analysis, in terms of SER, of an MP two-way massive MIMO system in presence of nonlinear PAs and running under a block fading Rayleigh channel. In this study, we considered a 3GPP PA model that reflects the realistic PA behavior. Analytical closed-form SER expressions have been developed giving insightful in the transmit power level needed as well as the deployment of MP two-way massive MIMO.
- 2) We investigate the potential of DL to enhance the energy-efficiency of MP two-way massive MIMO systems via the compensation of hardware imperfections (HWIs), which is very challenging. The HWIs are dominated by the power amplifiers (PA) adopted in the BS. Specifically, we propose a novel end-to-end (E2E) learning approach using agents [neural networks (NN)] within the different entities of the network, i.e., the user terminals and BS. Here, the optimization of the agents (i.e., find the NNs weights) is performed globally offline while each can be implemented independently in the inference mode, leading to a low computational complexity online. Interestingly, we conducted a comparative study showing the ability of our approach to provide better performance with lower computational complexity than the classical solutions proposed in literature.
- 3) It is worth mentioning that the success of our approach is related to the fact that the agent at the BS consists in a two-stage processing: the first stage aims at decoding the information and taking into account the PA nonlinearity, while the second stage consists in a linear precoding aiming to eliminate interuser interference. Note that by doing that we make a kind of channel charting that helps to make the training phase faster and more efficient as well as it makes the inference more efficient while requiring small NNs (i.e., with few weights to be optimized).

## D. Article Outline

The rest of this article is organized as follows: The MP two-way massive MIMO system in the presence of nonlinear PAs is introduced in Section II. Theoretical analysis of the performance in terms of SER of MP massive MIMO system, with and without, nonlinear PAs is presented in Section III. Section IV presents the proposed distributed learning technique. In Section V, numerical results and discussions are presented. Finally, Section VI concludes this article.

## E. Notation

Lower case boldface letters (e.g.,  $\mathbf{h}$ ) present a vectors, upper case boldface letters (e.g.,  $\mathbf{H}$ ) denote the matrix. The operators  $()^T$  and  $()^H$  the transpose and the conjugate transpose, respectively. The operators  $E[\cdot]$ ,  $|\cdot|$  and  $\|\cdot\|$  denote the expected value, the module, and the norm, respectively. The matrix  $\mathbf{I}$  denotes the identity matrix and the  $M_t \times M_r$  present the dimensional matrix.

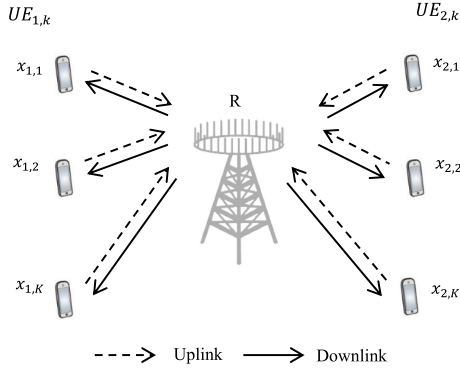


Fig. 1. Multi pair two-way massive MIMO system.

## II. MULTIPAIR TWO-WAY MASSIVE MIMO SYSTEM

In this investigation, we consider a multipair two-way massive MIMO system in which  $K$  pairs (equivalent to  $M_r = 2K$ ) of single-antenna devices communicate with each other in pairs, through a relay with  $M_t$ -antennas ( $M_t \gg M_r$ ). The relay and all devices operate in time-division duplex mode, so the channels between each device and the relay are reciprocal in each block. We consider block fading channel model, where the channels change independently from block to block while remaining constant during a block period.

### A. System Model

In the considered MP two-way half-duplex AF relaying system, the  $K$  pairs of single-antenna users, defined as  $UE_{1,k}$  and  $UE_{2,k}$ ,  $k = 1, \dots, K$ , exchange information between them in pairs and via a shared  $M_t$ -antenna relay (R), as shown in Fig. 1. Note that there are no direct transmission links between user pairs. Moreover, we assume a TDD mode which separates the downlink (DL) and on uplink (UL) transmissions with the assumption of perfect channel reciprocity, assured by accurate calibration methods [24]. Toward this end, the channel is assumed to be block Rayleigh fading one, which is constant during a time-frequency interval, known as the *coherence interval*, and varies independently between coherence intervals. The UL and DL channels between  $UE_{x,k}$ ,  $x = 1, 2$  and R are denoted by  $\mathbf{h}_{R_{x,k}} \sim \mathcal{CN}(0, \beta_{R_{x,k}} \mathbf{I}_M)$  and  $\mathbf{h}_{R_{x,k}}^T$ ,  $k = 1, \dots, K$ , respectively.  $\beta_{R_{x,k}}$  denotes the large-scale fading (variance of the channel), which is antenna independent. For the sake of simplicity and without loss of generality, we assume that the large-scale fading coefficients vary slowly, so it is constant over many coherence intervals.

### B. Uplink Transmission

The uplink channel matrix can be constructed as  $\mathbf{H}_{ux} = [\mathbf{h}_{1,1}, \dots, \mathbf{h}_{1,K}] \in \mathbb{C}^{M_t \times K}$ ,  $x = 1, 2$ , the downlink channel matrix can be computed as  $\mathbf{H}_d$ . In the first phase, all the devices  $UE_{1,k}$  and  $UE_{2,k}$  transmit simultaneously their signals carrying information, i.e.,  $x_{1,k}$  and  $x_{2,k}$ ,  $k = 1, \dots, K$  to relay R, respectively. We assume that  $x_{1,k}$  and  $x_{2,k}$  are QAM modulated data signals and that all devices transmit with the same power

(uplink), denoted by  $P_u$ . The received signal at R is

$$\mathbf{r}_u = \sum_{k=1}^K (\mathbf{h}_{1,k} x_{1,k} + \mathbf{h}_{2,k} x_{2,k}) + \mathbf{n}_u \quad (1)$$

where  $\mathbf{n}_u \in \mathbb{C}^{M_t \times 1}$  is an additive white Gaussian noise (AWGN uplink) with independent and identically distributed (i.i.d.) components with zero mean and  $\sigma_R^2$  variance. The expression (1) can also be written in a matrix form as

$$\mathbf{r}_u = \mathbf{H}_u \mathbf{x} + \mathbf{n}_u \quad (2)$$

where

- 1)  $\mathbf{H}_u = [\mathbf{H}_{u1}, \mathbf{H}_{u2}]$ ;
- 2)  $\mathbf{H}_{u1} = [\mathbf{h}_{u1,1}, \dots, \mathbf{h}_{u1,K}] \in \mathbb{C}^{M_t \times K}$ ;
- 3)  $\mathbf{H}_{u2} = [\mathbf{h}_{u2,1}, \dots, \mathbf{h}_{u2,K}] \in \mathbb{C}^{M_t \times K}$ ;
- 4) let  $\mathbf{H}_u \in \mathbb{C}^{M_t \times \overbrace{2K}^{M_r}}$ ;
- 5)  $\mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T]^T$ ;
- 6)  $\mathbf{x}_1 = [x_{1,1}, \dots, x_{1,K}]^T \in \mathbb{C}^{K \times 1}$ ;
- 7)  $\mathbf{x}_2 = [x_{2,1}, \dots, x_{2,K}]^T \in \mathbb{C}^{K \times 1}$ ;
- 8) let  $\mathbf{x} \in \mathbb{C}^{\overbrace{2K}^{M_r} \times 1}$ ;
- 9) let  $\mathbf{n}_u \in \mathbb{C}^{M_t \times 1}$  is an AWGN.

### C. Downlink Transmission and Precoding

In the second phase, R multiplies the received signal by a precoding matrix  $\mathbf{P} \in \mathbb{C}^{M_t \times M_t}$  and a power control coefficient  $\alpha$ . Then, R broadcasts to all the UEs signals as an expression of the form [25]

$$\mathbf{s}_t = \alpha \mathbf{P} \mathbf{r}_u = \alpha \mathbf{P} (\mathbf{H}_u \mathbf{x} + \mathbf{n}_u) \quad (3)$$

$$\alpha = \sqrt{\frac{P_r}{\mathcal{P}_u \|\mathbf{P} \mathbf{H}_u\|^2 + \sigma_R^2 \|\mathbf{P}\|^2}} \quad (4)$$

where  $\mathcal{P}_r$  is the power of the signal to be transmitted from the relay and  $\sigma_R^2$  is the variance of the noise ( $\mathbf{n}_u$ ). The precoding matrix  $\mathbf{P}$  is defined by the following expression:

$$\mathbf{P} = \mathbf{H}_d^T (\mathbf{H}_d \mathbf{H}_d^T)^{-1} (\mathbf{H}_u^T \mathbf{H}_u)^{-1} \quad (5)$$

where  $\mathbf{H}_d = [\mathbf{H}_{d1}, \mathbf{H}_{d2}]$ ,  $\mathbf{H}_{d1} = [\mathbf{h}_{d1,1}, \dots, \mathbf{h}_{d1,K}] \in \mathbb{C}^{K \times M_t}$ , and  $\mathbf{H}_{d2} = [\mathbf{h}_{d2,1}, \dots, \mathbf{h}_{d2,K}] \in \mathbb{C}^{K \times M_t}$  let  $\mathbf{H}_d \in \mathbb{C}^{\overbrace{2K}^{M_r} \times M_t}$ . The received signal by the devices on  $UE_{1,i}$  is

$$\begin{aligned} \mathbf{r}_d &= \mathbf{h}_{d1,K}^T \mathbf{s}_t + \mathbf{n}_d = \mathbf{h}_{d1,K}^T (\alpha \mathbf{P} \mathbf{r}_u) + \mathbf{n}_{x1,K} \\ &= \mathbf{h}_{d1,K}^T \left( \alpha \mathbf{P} \left( \sum_{k=1}^K (\mathbf{h}_{u1,k} x_{1,k} + \mathbf{h}_{u2,k} x_{2,k}) + \mathbf{n}_u \right) \right) + \mathbf{n}_{x1,k} \end{aligned} \quad (6)$$

where  $\mathbf{n}_d = \mathbf{n}_{x1,k} \in \mathbb{C}^{\overbrace{2K}^{M_r} \times 1}$  is an AWGN (downlink) with i.i.d  $\mathcal{N}$  components [25].

### D. Multipair Massive MIMO Under PA Nonlinearities

In this section, we present the effect of the nonlinear distortion due to PA on the MP massive MIMO system with  $M_t$  antennas



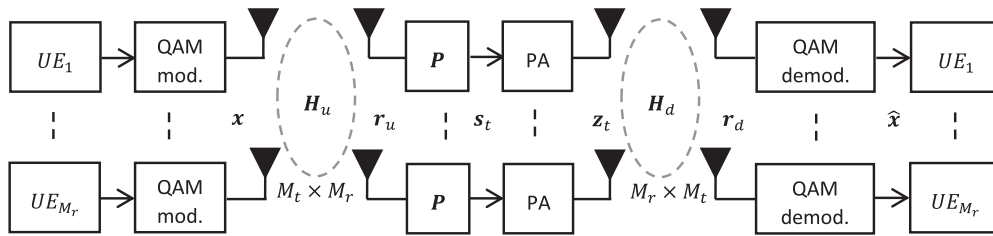


Fig. 2. Classical multi pair massive MIMO system under PAs nonlinearity.

and  $M_r$  users (pairs of  $K$  users) represented by Fig. 2. The recovered signals by the UEs ( $UE_{1,k}$ ) containing the received information from the relay will be amplified by nonlinear PAs without memories and subsequently retransmitted to the UEs ( $UE_{2,k}$ ). Neglecting memory can be justified, as we are primarily investigating how nonlinear distortion is processed over an MP massive MIMO downlink system, with high energy efficiency. The PAs are characterized by their dynamic amplitude distortion AM/AM and phase distortion AM/PM, which depend only on the current value of the input signal. Considering the complex envelope  $\mathbf{s}_t = \rho e^{j\theta}$  the input PA at each branch, where  $\rho$  and  $\theta$  denote, respectively, the amplitude and the phase of  $\mathbf{s}_t$ . The recovered signals at the output of the amplifiers without memory  $\mathbf{z}_t$  can be modeled as follows [26]:

$$\mathbf{z}_t = \mathcal{F}(\mathbf{s}_t) = \mathcal{F}_A(\beta\rho_k)e^{j(\theta_k + \mathcal{F}_P(\beta\rho_k))} \quad (7)$$

where  $\mathcal{F}(\cdot)$  is the HPA transfer function,  $\mathcal{F}_A(\cdot)$  and  $\mathcal{F}_P(\cdot)$  denote the PA amplitude conversion (AM/AM) and phase conversion (AM/PM), respectively. The factor  $\beta$  is a multiplicative coefficient added at the input of the PA for an operating point with a given input back-off (IBO). Here, all IBOs are given in dB relatives to the back-off from 1-dB compression point. The factor  $\delta$  necessary to ensure a signal  $\mathbf{s}_t$  with a given IBO value is  $\beta = \sqrt{\frac{P_{1dB}}{10 \frac{P_{1dB}}{IBO} P_t}}$ , where  $P_t$  is the transmit power [26].

In literature, several PAs have been proposed to amplify a transmitted signal. In this work, we will consider the modified Rapp model which fits closely to realistic PAs.

### E. SSPA Model

In the case of modified Rapp, the AM/AM and AM/PM conversions can be represented by

$$\mathcal{F}_A(|\mathbf{s}_t(n)|) = \frac{G|\mathbf{s}_t(n)|}{\left(1 + \left(\frac{G|\mathbf{s}_t(n)|}{A_s}\right)^{2p}\right)^{\frac{1}{2p}}} \quad (8)$$

$$\mathcal{F}_P(|\mathbf{s}_t(n)|) = \frac{A|\mathbf{s}_t(n)|^q}{\left(1 + \left(\frac{|\mathbf{s}_t(n)|}{B}\right)^q\right)} \quad (9)$$

where  $\mathbf{s}_t$  represents the amplitude of the input signal,  $G$  is the small signal gain,  $A_s$  is the saturation level,  $p$  is the softness factor, and  $A$ ,  $B$ , and  $q$ , respectively, represent the adjustment parameters [26].

On the other hand, since the signal at the input of the PA is Gaussian (massive MIMO precoding with large number of antennas) we can interpret the amplified signal by using the theorem of Bussgang [27] and the equivalent expression of the amplified signal can be written in the following form:

$$\mathbf{z}_t = \mathbf{W}\mathbf{s}_t + \mathbf{d} \quad (10)$$

where  $\mathbf{W} = \text{diag}([w_1, w_2, \dots, w_{M_t}])$  that is a square diagonal matrix with  $\{w_{m_t}\}$  elements (complex gain) on the diagonal and  $\mathbf{d} \sim \mathcal{CN}(0, \sigma_d^2 \mathbf{I}_{M_t})$  represents the Gaussian distortion noise with zero mean and variance  $\sigma_d^2$ . It should be noted that all PAs work with the same IBO to balance the power consumption between the different branches in a massive MIMO. More precisely, by means of a mathematical derivation [28], these NLD parameters ( $w_{m_t}$  and  $\sigma_d^2$ ) can be calculated analytically for any PAs.

Then, we can write the amplified signal at each branch as  $z_{m_t} = w_{m_t}s_{m_t} + d_{m_t}$ , where  $w_{m_t}$  is the  $m_t$ th complex gain of PA and  $d_{m_t}$  represents the Gaussian distortion noise.

Finally, the input-output relationship of the MP massive MIMO system with nonlinear PAs can be expressed as

$$\mathbf{r}_d^{NL} = \mathbf{H}_d \mathbf{z}_t + \mathbf{n}_d \quad (11)$$

where  $\mathbf{n}_d \sim \mathcal{CN}(0, \sigma_{n_d}^2)$  is AWGN.

By replacing (10) in (11), we obtain

$$\mathbf{r}_d^{NL} = \mathbf{H}_d \mathbf{W} \mathbf{s}_t + \mathbf{H}_d \mathbf{d} + \mathbf{n}_d. \quad (12)$$

### III. THEORETICAL ANALYSIS

In this section, we consider a half-duplex AF relaying as classical method for which an analytical study has been conducted to evaluate the impact of the nonlinear PAs on the (SER) performance of MP two-way massive MIMO.

In [29], [30], and [31], the performance analysis in terms of bit error rate (BER) and (SER) is well studied. A fading channel is considered as an AWGN with a variable gain, and the gain itself is examined as a random variable with a given probability density function (pdf).

Moreover, the average SER can be calculated by averaging SER for instantaneous signal-to-interference-plus-noise ratio (SINR) over its distribution, using the quadrature amplitude modulation (M-QAM)

$$\text{SER} = \int_0^{+\infty} \text{SER}_{\text{M-QAM}} p_\gamma(\gamma) d\gamma \quad (13)$$

where  $p_\gamma(\gamma)$  is the pdf of the instantaneous SINR( $\gamma$ ).

So, the SER of a system, which implements M-QAM over a Rayleigh fading channel, can be defined as [30]

$$\text{SER}_{\text{M-QAM}} = 4Q \left( \sqrt{\frac{3\gamma \log_2(M)}{M-1}} \right) \quad (14)$$

where  $M$  is the modulation order and  $Q$  is the standard Q-function [30].

We develop a new expression of the (SER) for the MP massive MIMO system over Rayleigh fading channel with and without

power amplifier. Thus, we will establish the average SER equation, and determine the SINR( $\gamma$ ) equation and its pdf ( $p_\gamma(\gamma)$ ) for each scenario.

#### A. Performance Analysis of MP Massive MIMO System Over Linear Rayleigh Fading Channel

Using (6), the received signal by the devices on  $UE_{(1,k)}$  can be written as

$$\begin{aligned} \mathbf{r}_d &= \mathbf{h}_{d1,k}^T \left( \alpha \mathbf{P} \left( \sum_{k=1}^K (\mathbf{h}_{u1,k} x_{1,k} + \mathbf{h}_{u2,k} x_{2,k}) + \mathbf{n}_u \right) \right) + \mathbf{n}_d \\ &= \underbrace{\alpha \mathbf{h}_{d1,k}^T \mathbf{P} \mathbf{h}_{u2,k} x_{2,k}}_{\text{signal}} + \underbrace{\alpha \mathbf{h}_{d1,k}^T \mathbf{P} \mathbf{h}_{u1,k} x_{1,k}}_{\text{selfinterference}} \\ &\quad + \underbrace{\alpha \mathbf{h}_{d1,k}^T \sum_{j=1, j \neq k}^K (\mathbf{P} \mathbf{h}_{u1,j} x_{1,j} + \mathbf{P} \mathbf{h}_{u2,j} x_{2,j})}_{\text{interuserinterference}} \\ &\quad + \underbrace{\alpha \mathbf{h}_{d1,k}^T \mathbf{P} \mathbf{n}_u + \mathbf{n}_d}_{\text{compoundnoise}}. \end{aligned} \quad (15)$$

To determine the SINR equations, we will consider  $\mathbf{h}_{u1} = \mathbf{h}_{u2} = \mathbf{h}_u$  and  $\mathbf{h}_{d1} = \mathbf{h}_{d2} = \mathbf{h}_d$ .

$$\begin{aligned} \mathbf{r}_d &= \mathbf{h}_{d,k}^T \left( \alpha \mathbf{P} \left( \sum_{k=1}^K (\mathbf{h}_{u,k} x_{1,k} + \mathbf{h}_{u,k} x_{2,k}) + \mathbf{n}_u \right) \right) + \mathbf{n}_d \\ &= \alpha \mathbf{h}_{d,k}^T \mathbf{P} \mathbf{h}_{u,k} x_{2,k} + \alpha \mathbf{h}_{d,k}^T \mathbf{P} \mathbf{h}_{u,k} x_{1,k} \\ &\quad + \alpha \mathbf{h}_{d,k}^T \sum_{j=1, j \neq k}^K (\mathbf{P} \mathbf{h}_{u,j} x_{1,j} + \mathbf{P} \mathbf{h}_{u,j} x_{2,j}) \\ &\quad + \alpha \mathbf{h}_{d,k}^T \mathbf{P} \mathbf{n}_u + \mathbf{n}_d. \end{aligned} \quad (16)$$

The SINR of  $UE_{1,k}$  can be expressed as

$$\gamma_{x_{1,k}} = \frac{|\alpha|^2 \mathcal{P}_u |\mathbf{h}_{d,k}^T|^2 |\mathbf{P} \mathbf{h}_{u,k}|^2}{\mathcal{A} + \mathcal{B} + \mathcal{C} + \mathcal{D}} \quad (17)$$

where

- 1)  $\mathcal{A} = |\alpha|^2 \mathcal{P}_u |\mathbf{h}_{d,k}^T|^2 \sum_{j=1, j \neq k}^K (|\mathbf{P} \mathbf{h}_{u,j}|^2 + |\mathbf{P} \mathbf{h}_{u,j}|^2)$ ;
- 2)  $\mathcal{B} = |\alpha|^2 \mathcal{P}_u |\mathbf{h}_{d,k}^T|^2 |\mathbf{P} \mathbf{h}_{u,k}|^2$ ;
- 3)  $\mathcal{C} = |\alpha|^2 \sigma_R^2 |\mathbf{h}_{d,k}^T|^2 |\mathbf{P}|^2$ ;
- 4)  $\mathcal{D} = \sigma_{nd}^2$  and  $|\mathbf{h}_{d,k}^T|^2$  is assumed to be a flat fading one, and it is exponentially distributed with a pdf given by  $\mathcal{P}_\gamma(\gamma) = \frac{1}{\Omega} \exp^{-\frac{\gamma}{\Omega}}$ , where  $\Omega = \mathcal{E}[|\mathbf{h}_{d,k}^T|^2]$  [29].

The average SER of the M-QAM MP massive MIMO system without PAs can be expressed as

$$\text{SER} = \frac{M-1}{M} \left( 1 - \frac{1}{\sqrt{1 + \frac{M^2-1}{3\gamma \log_2 M}}} \right). \quad (18)$$

#### B. Performance Analysis of Nonlinear MP Massive MIMO Over Rayleigh Fading Channel

According to (11) and (12) on the one hand and Fig. 2 on the other hand, the signal recovered at the output of the power

amplifier is given as follows:

$$\begin{aligned} \mathbf{r}_d^{NL} &= \mathbf{h}_{d1,k} (\mathbf{W} \mathbf{s}_t + \mathbf{d}) + \mathbf{n}_d \\ &= \mathbf{h}_{d1,k} \left( \mathbf{W} \left( \alpha \mathbf{P} \left( \sum_{k=1}^K (\mathbf{h}_{u1,k} x_{1,k} + \mathbf{h}_{u2,k} x_{2,k}) \right) \right) \right. \\ &\quad \left. + \mathbf{n}_u \right) + \mathbf{d} + \mathbf{n}_d \\ &= \mathbf{h}_{d1,k} \left( \mathbf{W} \alpha \mathbf{P} \sum_{k=1}^K (\mathbf{h}_{u1,k} x_{1,k} + \mathbf{h}_{u2,k} x_{2,k}) \right. \\ &\quad \left. + \mathbf{W} \alpha \mathbf{P} \mathbf{n}_u + \mathbf{d} \right) + \mathbf{n}_d \\ &= \mathbf{W} \alpha \mathbf{h}_{d1,k}^T \mathbf{P} \sum_{k=1}^K (\mathbf{h}_{u1,k} x_{1,k} + \mathbf{h}_{u2,k} x_{2,k}) \\ &\quad + \mathbf{W} \alpha \mathbf{h}_{d1,k}^T \mathbf{P} \mathbf{n}_u + \mathbf{h}_{d1,k}^T \mathbf{d} + \mathbf{n}_d. \end{aligned} \quad (19)$$

Considering  $\mathbf{h}_{u1} = \mathbf{h}_{u2} = \mathbf{h}_u$  and  $\mathbf{h}_{d1} = \mathbf{h}_{d2} = \mathbf{h}_d$  the amplified signal received by the devices on  $UE_{1,k}$  can be interpreted as

$$\begin{aligned} \mathbf{r}_d^{NL} &= \underbrace{\mathbf{W} \alpha \mathbf{h}_{d,k}^T \mathbf{P} \mathbf{h}_{u,k} x_{2,k}}_{\text{amplifiedsignal}} + \underbrace{\mathbf{W} \alpha \mathbf{h}_{d,k}^T \mathbf{P} \mathbf{h}_{u,k} x_{1,k}}_{\text{selfinterference}} \\ &\quad + \underbrace{\mathbf{W} \alpha \sum_{j=1, j \neq k}^K \mathbf{h}_{d1,k}^T (\mathbf{P} \mathbf{h}_{u,j} x_{1,j} + \mathbf{P} \mathbf{h}_{u,j} x_{2,j})}_{\text{interuserinterference}} \\ &\quad + \underbrace{\mathbf{W} \alpha \mathbf{h}_{d,k}^T \mathbf{P} \mathbf{n}_u}_{\text{hardwareimperfections(PA)}} + \underbrace{\mathbf{h}_{d,k}^T \mathbf{d} + \mathbf{n}_d}_{\text{compoundnoise}}. \end{aligned} \quad (20)$$

The SINR<sup>NL</sup> of  $UE_{x_{1,k}}$  can be developed as follows:

$$\gamma_{x_{1,k}}^{NL} = \frac{|\mathbf{W}|^2 |\alpha|^2 \mathcal{P}_u |\mathbf{h}_{d,k}^T|^2 |\mathbf{P} \mathbf{h}_{u,k}|^2}{\mathcal{A}_1 + \mathcal{B}_1 + \mathcal{C}_1 + \mathcal{D}_1 + \mathcal{E}_1} \quad (21)$$

where

- 1)  $\mathcal{A}_1 = |\mathbf{W}|^2 |\alpha|^2 \mathcal{P}_u |\mathbf{h}_{d,k}^T|^2 2 \sum_{j=1, j \neq k}^K |\mathbf{P} \mathbf{h}_{u,k}|^2$ ;
- 2)  $\mathcal{B}_1 = |\mathbf{W}|^2 |\alpha|^2 \mathcal{P}_u |\mathbf{h}_{d,k}^T|^2 |\mathbf{P} \mathbf{h}_{u,k}|^2$ ;
- 3)  $\mathcal{C}_1 = |\mathbf{W}|^2 |\alpha|^2 \sigma_R^2 |\mathbf{h}_{d,k}^T|^2 |\mathbf{P}|^2$ ;
- 4)  $\mathcal{D}_1 = \sigma_d^2 |\mathbf{h}_{d,k}^T|^2$ ;
- 5)  $\mathcal{E}_1 = \sigma_{nd}^2$ .

Dividing equation (18) by  $|\mathbf{W}|^2 |\alpha|^2$ , we obtain

$$\gamma_{x_{1,k}}^{NL} = \frac{\mathcal{P}_u |\mathbf{h}_{d,k}^T|^2 |\mathbf{P} \mathbf{h}_{u,k}|^2}{\mathcal{A}'_1 + \mathcal{B}'_1 + \mathcal{C}'_1 + \mathcal{D}'_1 + \mathcal{E}'_1} \quad (22)$$

where

- 1)  $\mathcal{A}'_1 = \mathcal{P}_u |\mathbf{h}_{d,k}^T|^2 2 \sum_{j=1, j \neq k}^K |\mathbf{P} \mathbf{h}_{u,j}|^2$ ;
- 2)  $\mathcal{B}'_1 = \mathcal{P}_u |\mathbf{h}_{d,k}^T|^2 |\mathbf{P} \mathbf{h}_{u,k}|^2$ ;
- 3)  $\mathcal{D}'_1 = \frac{\sigma_d^2}{|\mathbf{W}|^2 |\alpha|^2} |\mathbf{h}_{d,k}^T|^2$ ;
- 4)  $\mathcal{E}'_1 = \frac{\sigma_{nd}^2}{|\mathbf{W}|^2 |\alpha|^2}$ .

Let

- 1)  $\Psi = |\mathbf{h}_{d,k}^T|^2$ ;

- 2)  $\gamma_h = |\mathbf{P}\mathbf{h}_{u,k}|^2$ ;
- 3)  $\gamma_{h,j} = \sum_{j \neq k}^K |\mathbf{P}\mathbf{h}_{u,j}|^2$ ;
- 4)  $\gamma_{NL} = \frac{\sigma_d^2}{|\mathbf{W}|^2|\alpha|^2}$ ;
- 5)  $\gamma_n = \sigma_R^2 + \frac{\sigma_d^2}{|\mathbf{W}|^2|\alpha|^2}$ , equation (22) becomes

$$\gamma_{x1,k}^{NL} = \frac{\mathcal{P}_u \Psi \gamma_h}{\mathcal{P}_u \Psi \gamma_h + 2\mathcal{P}_u \Psi \gamma_{h,j} + \gamma_{NL} \Psi + \gamma_n} = F(\Psi) \quad (23)$$

and the pdf is given by  $p_{\gamma_{x1,k}^{NL}}(\gamma) = p_{\Psi}(g(\gamma))g'(\gamma)$ . The  $\gamma_{x1,k}^{NL}$  can be expressed by

$$\gamma = \frac{\mathcal{P}_u \gamma_h g(\gamma)}{\mathcal{P}_u \gamma_h g(\gamma) + 2\mathcal{P}_u \gamma_{h,j} g(\gamma) + \gamma_{NL} g(\gamma) + \gamma_n}. \quad (24)$$

From (24), we can establish  $g(\gamma)$  and the derivative of  $g(\gamma)$  ( $g'(\gamma)$ )

$$g(\gamma) = \frac{\gamma_n \gamma}{\mathcal{P}_u \gamma_h - \mathcal{P}_u \gamma_h \gamma - 2\mathcal{P}_u \gamma_{h,j} \gamma - \gamma_{NL} \gamma} \quad (25)$$

$$g'(\gamma) = \frac{\mathcal{P}_u \gamma_n \gamma_h}{(\mathcal{P}_u \gamma_h - \mathcal{P}_u \gamma_h \gamma - 2\mathcal{P}_u \gamma_{h,j} \gamma - \gamma_{NL} \gamma)^2}. \quad (26)$$

Thus, the average SER of the M-QAM nonlinear MP massive MIMO system is given by

$$\begin{aligned} \text{SER}^{NL} &= \int_0^{\frac{\gamma_h}{\gamma_h + 2\gamma_{h,j} + \frac{\gamma_{NL}}{\mathcal{P}_u}}} 4Q \left( \sqrt{\frac{3\gamma \log_2 M}{M-1}} \right) \\ &\quad \times \frac{\mathcal{P}_u \gamma_n \gamma_h}{\Omega(\mathcal{P}_u \gamma_h - \mathcal{P}_u \gamma_h \gamma - 2\mathcal{P}_u \gamma_{h,j} \gamma - \gamma_{NL} \gamma)^2} \\ &\quad \times \exp \left( -\frac{\gamma_n \gamma}{\mathcal{P}_u \gamma_h - \mathcal{P}_u \gamma_h \gamma - 2\mathcal{P}_u \gamma_{h,j} \gamma - \gamma_{NL} \gamma} \right). \end{aligned} \quad (27)$$

#### IV. DEEP LEARNING TECHNIQUE FOR NONLINEAR DISTORTION CANCELLATION

There is a strong demand for low-cost nonlinear PAs which, however, introduce nonlinear distortions. This is because modern and expensive APs exhibit high energy efficiency in the nonlinear region. We have tried to find the best operating point to ensure a good compromise between nonlinear distortion and efficiency.

On the other hand, we are still looking to deal with random data, support vector machines, logistic regression, etc., but traditional machine learning algorithms are generally inefficient as the size of the data increases. This problem is solved by using the advent of DL. It will allow us to train an AI to predict outcomes, based on a set of input data. In this section, we will use two compensation methods-based DL to eliminate a nonlinear distortion caused by PA in massive MIMO system.

##### A. Neural Networks-Based Compensation

In our work, we consider an architecture based on an indirect learning NN to process data using feed-forward technology, characterized by the passage of data from the input of the network to its output without returning in back. For this type of network, we used the multilayer network called the multilayer perceptron (MLP); Fig. 3 shows an example of its architecture. The basic idea of the proposed compensation technique is to

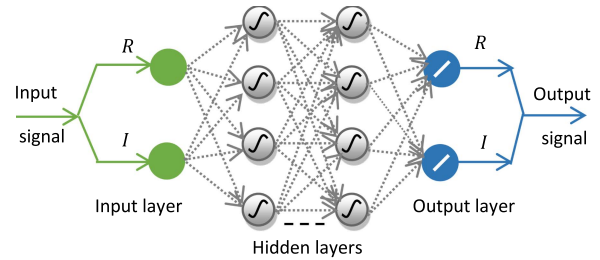


Fig. 3. Architecture of MLP network, where the indices R and I refer to the real and imaginary parts, respectively.

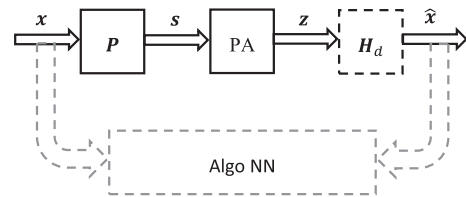


Fig. 4. Algo NN for MP massive system MIMO in the presence of PAs under Rayleigh fading channel.

identify the inverse PA transfer function using MLP networks. At this level, we will interpret an algorithm based on NNs (Algo NN) for the joint compensation of the nonlinear distortion caused by the PAs and channel effects on the massive MP MIMO system, as shown in Fig. 4.

1) *Multilayer Perceptron*: This type of network consists of an input layer, an output layer, and several hidden layers between these two layers, and the number of layers used corresponds to the number of weight matrices available. Thus, we can notice that a PMC is in this case considered as the network best adapted to treat nonlinear functions. Fig. 3 presents the architecture of the PMC having several layers, where, we can note that the first layer is connected directly to the inputs, each layer is connected to the previous and the next and the last layer gives the outputs. We must consider an adaptive NN, where the modifications of each parameter carried out on each neuron will be able to adapt the network in order to establish and carry out all types of operations, and these modifications will be interpreted and used during the learning phase [26].

2) *Compensation Algorithm Based on NN for MP Massive MIMO in the Presence of PAs*: The compensation algorithm-based NN (Algo NN) for the effects of PAs introduced in Fig. 4 aims to eliminate the effects of nonlinearities and establish an opposite response to that given by PAs (linear response). Thus, an additional nonlinear function is necessary to preprocess the input signal PA, and thus, linearize the amplified signal. The goal of our approach is to formulate a simple convex optimization method that allows to implement an efficient, simple, and reliable algorithm. For the nonlinear MP massive MIMO system, as introduced in this work, we can interpret the problem by minimizing the minimum square error (MSE) between the signal  $\mathbf{x}$  and the amplified signal  $\hat{\mathbf{x}}$  through the channel  $\mathbf{H}_d$  as

$$\text{MSE} = \|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 \quad (28)$$

$$\text{MSE} = \arg \min_{\mathbf{s}} \|\mathbf{H}_d f(\mathbf{s}) - \mathbf{x}\|_2^2. \quad (29)$$

It is difficult to directly address the problem of (29) because of the nonlinear operation induced by the PAs, where we can note the input signals of the PA are Gaussian and complex. Using the Bussgang's theorem given by [28], we can decompose the nonlinear signal recovered at the output of PA into a linear function and an uncorrelated distortion term. Alternatively, the iterative algorithm could achieve a very high-performance gain with insignificant computational complexity. By minimizing the mean square error (MSE) of the symbols  $\mathbf{x}$  and the output signal of PAs outputs given by the Bussgang's theorem  $\mathbf{z}$  through the  $\mathbf{H}_d$  channel, we can rewrite (29) as

$$\underset{\mathbf{s}}{\text{minimize}} \mathcal{J}(\mathbf{s}) = \|\mathbf{H}_d \mathbf{s} + \mathbf{H}_d \mathbf{d} - \mathbf{x}\|_2^2. \quad (30)$$

The proposed equation (12) will allow an efficient iterative NL distortion attenuation algorithm, which aims to search for  $s_m t$  symbols guaranteeing excellent transmission quality. Thus, we generate a local copy of the distortion vector received by the targeted users on the transmitter side, so that it can be used to adapt the precoded vector  $\mathbf{s}$  and would be canceled when the latter is amplified and passed through the channel. To calculate the distortion vector, we use an initial estimate of the vector  $\mathbf{s}^0$  using the direct approximate models PAs which introduce the NL distortion on the signal (10), in the same way as the real PAs in the real transmission, i.e., expressed by  $\mathbf{d} = F(\mathbf{s}) - \mathbf{W}\mathbf{s}$ .

Then, the channel estimates  $\mathbf{H}_d$  reproduce the distortion vector  $\mathbf{d} = \mathbf{H}_d \mathbf{d}$  at the receiver level emulating the true propagation. This estimated distortion vector is considered, and a new estimate of the vector  $\mathbf{s}$  can be obtained. This second estimate  $\mathbf{s}^1$  can be used to reestimate the distortion vector. This process can be performed iteratively until a certain specific limit is reached. Therefore, an alternative estimate has been presented in (12), in which we estimate  $\mathbf{s}$  by minimizing the function  $\mathcal{J}(\mathbf{s})$  with respect to  $\mathbf{s}$  and estimating the distortion vector. Thus, the alternative procedure in the  $(t+1)$ th iteration is expressed as

$$\mathbf{s}^{(t+1)} = \arg \min_{\mathbf{s}} (\mathcal{J}(\mathbf{s}^{(t)}, \mathbf{d}^{(t)})) \quad (31)$$

$$\mathbf{d}^{(t+1)} = F(\mathbf{s}^{(t+1)}) - \mathbf{W}\mathbf{s}^{(t+1)}. \quad (32)$$

So, the considered optimization problem lends itself to an efficient implementation for massive MIMO systems by offering an algorithm of low complexity and which only requires matrix-vector multiplications. The gradient descent method (GD) at iteration  $\mathbf{s}^{(t+1)}$  are determined by the negative gradient of the function cost  $\mathcal{J}$  to  $\mathbf{s}^{(t)}$ , which is given by

$$\nabla_{\mathbf{s}} \mathcal{J}(\mathbf{s}^{(t)}, \mathbf{d}^{(t)}) = 2\mathbf{W}^H \mathbf{H}_d^H [\mathbf{H}_d \mathbf{W}\mathbf{s}^{(t)} + \mathbf{H}_d \mathbf{d}^{(t)} - \mathbf{x}]. \quad (33)$$

To improve the rate of convergence of our algorithm, we add a coefficient  $\mu$  to ensure the link between the previous iteration and the current iteration. The vector  $\mathbf{s}$  will be adjusted according to

$$\mathbf{s}^{(t+1)} = \mathbf{s}^{(t)} - \Delta \mathbf{s}^{(t)} \quad (34)$$

where  $\Delta \mathbf{s}^{(t)} = \gamma \nabla_{\mathbf{s}} \mathcal{J}(\mathbf{s}^{(t)}, \mathbf{d}^{(t)}) - \mu \Delta \mathbf{s}^{(t-1)}$ . The proposed algorithm in the form of Algo NN is defined as follows.

### B. End-to-End Learning for MP Massive MIMO Systems in the Presence of PAs Using Autoencoder

In this section, we proposed an E2E wireless communication system based on the deep neural network (DNN) autoencoder to

---

#### Algorithm 1: Algo NN.

---

INITIALIZATION:

$\mathbf{s}^0 = \mathbf{0}_{M_t \times 1}$ ,  $\mathbf{d}^1 = \mathbf{0}_{M_r \times 1}$ ,  $\Delta \mathbf{s}^0 = \mathbf{0}_{M_t \times 1}$ , we fix the maximal number of iterations as  $T$ , the learning rate as  $\gamma$  and the momentum coefficient as  $\mu$

**for**  $t=1, \dots, T$  **do**

$\Delta \mathbf{s}^{(t)} =$

$2\gamma \mathbf{W}^H \mathbf{H}_d^H [\mathbf{H}_d \mathbf{W}\mathbf{s}^{(t)} + \mathbf{H}_d \mathbf{d}^{(t)} - \mathbf{x}] - \mu \Delta \mathbf{s}^{(t-1)}$

$\mathbf{s}^{(t+1)} = \mathbf{s}^{(t)} - \Delta \mathbf{s}^{(t)}$

Adjustment of power of  $\mathbf{s}^{(t+1)}$  to considered *ibo*.

$\mathbf{d}^{(t+1)} = F(\mathbf{s}^{(t+1)}) - \mathbf{W}\mathbf{s}^{(t+1)}$

**end for**

**return**  $\mathbf{s}^{(T+1)}$

---

eliminate jointly the effect of power amplifier and imperfect channel state information for MU massive MIMO systems. The DNNs are applied to perform existing functions in the transmission chain, such as modulation, demodulation, coding, decoding, and PA, etc.

Indeed, autoencoders are unsupervised learning algorithms based on artificial NNs that can generate new representations of datasets, which are generally more compact and have fewer descriptors, thus reducing the size of the dataset. The architecture of the autoencoder consists of two parts: 1) an encoder and 2) a decoder.

- 1) *The encoders*: They consist of a set of layers of neurons that process the data to build new representations called "encoded."
- 2) *The decoders*: The layers of neurons in the decoder receive these representations and process them to reconstruct the original data with a minimal deviation.

In our work, we will consider the MP massive MIMO system, where the relay has a large number of  $M_t$  antennas with PAs and aims to exchange the symbols between several pairs users ( $M_r = 2K$ ). By exploiting the concept of the autoencoder, in which the E2E transmission system is modeled using a DNN. Fig. 5 gives us an idea of the working principle of our proposed technique. This adopted diagram shows us how to use a nonlinear amplifier as an autoencoder to represent an E2E communication system and how to train this autoencoder to send the symbols and eliminate the effects of PAs and the channel.

However, the main goal of the transmission system under consideration is to successfully return the information sent to the end users. This communication task can be approached as a classification problem for which categorical cross-entropy is a frequent choice of loss function. We recall that the autoencoder used can be trained E2E through gradient descent algorithms to minimize the average cross-entropy between the one-hot representation of the predicted messages and the probability vectors generated by the user [32]. Specifically, the cross-entropy constrains the output of the decoder to be the same symbol as the input to the precoder. The gradients are back-propagated through the decoder and then the encoder, which allows us to perform weight updates.

At this level, we emphasize that a common DNN is used to present the encoding and decoding procedures of different users or each one is modeled with two hidden layers and an activation function (Relu). They are trained with 40 epochs (in other words, 40 different channels), chosen at random.



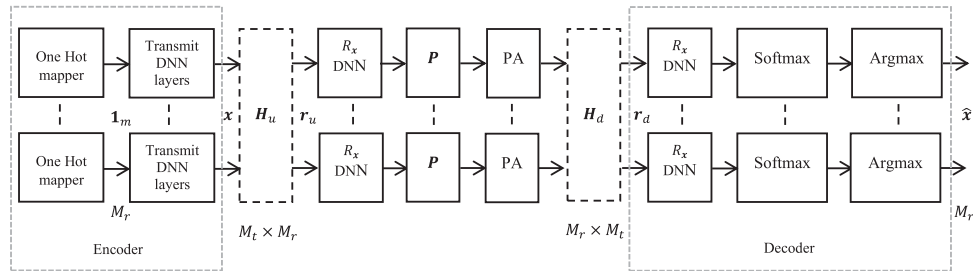


Fig. 5. End-to-end learning architecture for a nonlinear MP massive MIMO system.

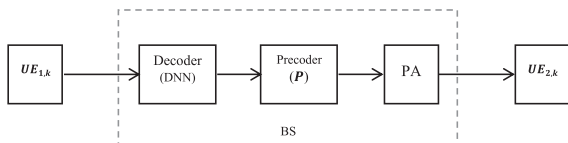


Fig. 6. Relay-level processing.

TABLE I  
PA PARAMETERS

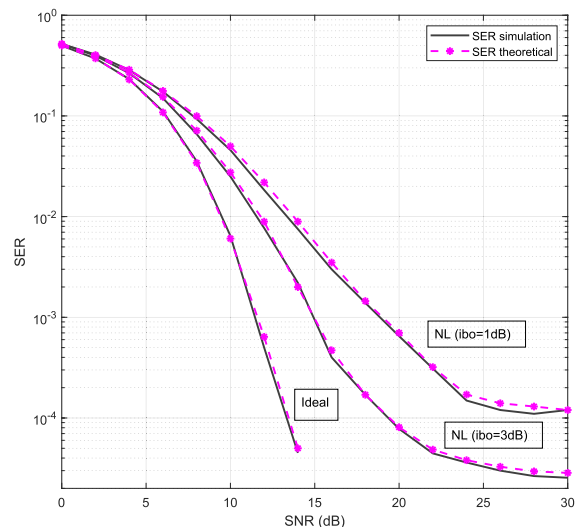
PA	Parameters
SSPA	$G = 16, A_s = 1.9, p = 1.1, A = -345, B = 0.17, q = 4$
ibo	1 dB, 3 dB

To begin with, most existing DL libraries only support the manipulation of real values, so we must first transform the complex model into a real model by concatenating the real and imaginary parts.

Thus, we consider  $1_m$  as an autoencoder input containing information sent to the user. This vector is transferred to the transmission DNN which has several nonlinear layers followed by a linear layer. This information will be transmitted via an uplink channel. In order to eliminate the effects of the channel and to decode the information received at the relay level (R), we implemented a decoder with several layers. At this level, R multiplies the decoded signal by a precoding matrix  $P$ . Then, the obtained vector will be amplified by nonlinear APs. Finally, this vector is broadcast to all users via a downlink channel (as shown in Fig. 6). At the receiver, another DNN with multiple layers is considered as a decoder, as in the case of the transmitter encoder. At the receiving DNN, the activation function of the last layer is defined as a softmax activation function to generate the probability vector at user equipment  $k$  [23].

## V. SIMULATIONS RESULTS

In this section, we present the simulation results and performance theoretical analysis in terms of SER of a nonlinear multipair massive MIMO system under Rayleigh fading channel with QAM modulation. We plot the SER as a function of the SNR of the two approaches (algo NN and EA), where the SNR is defined as  $\text{SNR} = \frac{E\|x\|_2^2}{N_0}$  and the parameters of the considered amplifiers are shown in Table I. At this level, we consider random data of size  $10^5$ . These data are processed and then stored to facilitate learning.

Fig. 7. Comparison of simulation and theoretical SER in nonlinear (NL) MP massive MIMO system for  $M_r = 10$ .

### A. Comparison Between Theoretical and Simulation Results

The theoretical and simulated average SER performances of the MP massive MIMO system with and without PA are shown in Fig. 7. The theoretical curves are obtained from the theoretical expression [(18) and (26)] of the SER and are given by marker lines. The simulated curves are represented solid lines. The theoretical values of the nonlinear parameters, such as the complex gain  $w$  and the variance  $\sigma_d^2$  of the nonlinear noise  $d(t)$  are interpreted and calculated in [28] and [29]. Fig. 7 shows that the theoretical results are very close to the simulated ones. On the other hand, it presents the effect of the power amplifier on the performance of the considered system especially for a low value of  $ibo$  (1 dB). On the other hand, one can notice the effect of the PAs for SNR values higher than 25 dB (constant level).

### B. Performance of Compensation Techniques Based on Deep Learning

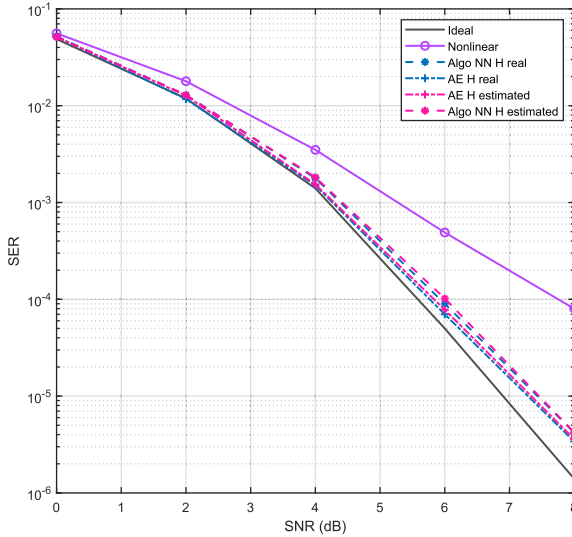
Indeed, several methods have been proposed to eliminate imperfections caused by PAs, such as: iterative algorithms [33], PAPR reduction, and DPD [19] on different transmission systems. In this context, we first prepared a comparative table (Table II), where we compared our two proposed methods with those used in the literature. Second, we will introduce our two methods proposed in this article: an algorithm based on NNs (Algo NN) and a new E2E (AE) approach allowing to establish

TABLE II  
 PERFORMANCE COMPARISON ( $M_r = 20$  AND  $ibo=3$ )

Techniques	SER for SNR = 10 dB	SER for SNR = 14 dB
DPD	0.10905	0.0095
PAPR reduction	0.10334	0.00797
Algo NN	0.10306	0.00782
AE	0.0094	0.0051

 TABLE III  
 SIMULATION PARAMETERS

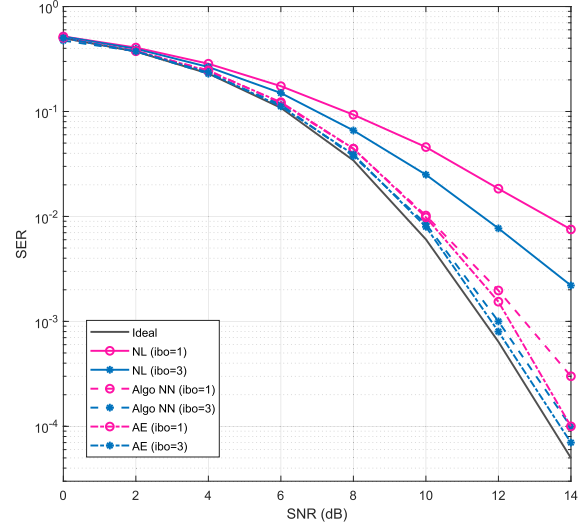
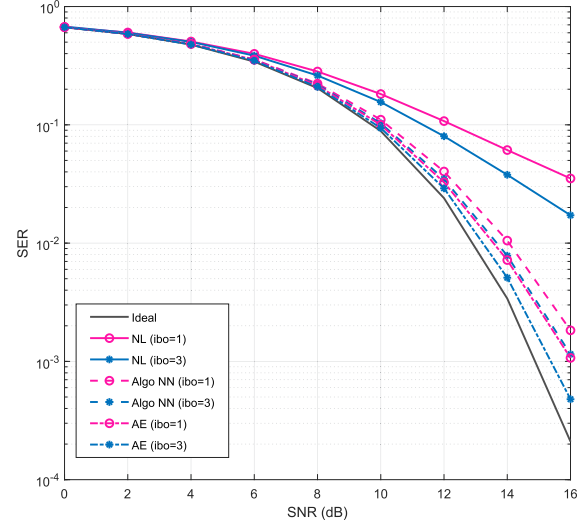
Parameters	Values
Number of users ( $M_r$ )	2, 10, 20
Number of antennas ( $M_t$ )	100
QAM order	16


 Fig. 8. Comparison of Algo NN and AE in MP massive MIMO system for  $M_r = 2$  and  $ibo = 3$  dB for different types of channels.

the same transmitter–receiver transmission chain of the MP massive MIMO system. When coming to the proposed approach, the NN adopted at the BS has as input the received signal, i.e., at the channel output. This NN has as role to perform the first stage precoding and the generated signal will serve as an input to the linear precoder (i.e., second stage precoding). This NN has many layers (i.e., deep) can be seen as a black box but it can perform like a decoder and precoder (it is no longer a classical AF), leading to an improved performance compared to classical methods. The simulation parameters for these solutions are shown in Table III.

Fig. 8 presents the impact of the channel estimation error on the SER performance. We consider the channel matrix in the frequency domain as represented by:  $\hat{\mathbf{H}}_d = \mathbf{H}_d + \tilde{\mathbf{H}}_d$ , where  $\tilde{\mathbf{H}}_d \sim \mathcal{CN}(0, \sigma_{\tilde{h}_d}^2)$  is the channel estimation error whose entries are complex Gaussian distributed with a variance of  $\sigma_{\tilde{h}_d}^2$  [18].

Figs. 8–10 show the performance of the iterative algorithm Algo NN and the AE approach for different numbers of users (where  $M_r = 2K$ ). Whatever the number of users, by analyzing the curves given by the figures the nonlinear distortion can be considerably compensated by using the two proposed techniques, except that, the curves given by the AE method tend to approach at most the linear curve.


 Fig. 9. Comparison of Algo NN and AE in MP massive MIMO system for  $M_r = 10$ .

 Fig. 10. Comparison of Algo NN and AE in MP massive MIMO system for  $M_r = 20$ .

### C. Complexity Analysis

In this work, we develop a complexity study focused on the algo NN based on the multilayer perceptron and the E2E method-based DNN. The Algo NN is applied to each branch of antenna before the power amplifier. More precisely, we considered for this algo a single layer, one hidden layer with  $N_n$  neurons. Most learning libraries accept only real values, so, we transformed the complex model into a real model (two neurons in the input layer and two neurons in the output layer). In the fact, to compensate the effect of PA nonlinearities, we considered the Algo NN learning off-line, then, their complexity is not considered. According to the proposed algorithm, we obtained the complexity given by  $\mathcal{O}(T(3(M_t M_r + M_t)))$ , where  $T$  presents the maximum number of iterations. To simplify the complexity of the proposed algorithms to eliminate the hardware imperfections, we have interpreted the E2E approach. The advantage of this approach is that all blocks in the chain are processed offline.

## VI. CONCLUSION

In this article, we presented a theoretical analysis for the SER performance of a multipair two-way massive MIMO system, in presence of non linear PA and under independent Rayleigh block fading channel. This analysis can provide important insights into the practical impact of power-efficient PA and also how to efficiently deploy MP two-way massive MIMO systems. Furthermore, we presented two DL-based techniques to eliminate the nonlinear distortion caused by PA and then enhance the energy efficiency of the system. The first proposed method consists in an algorithm based on an NN (Algo NN), it is designed with low complexity architecture, enabling energy-efficient MP two-way massive MIMO. Interestingly, the second is a novel E2E learning approach, whereas the E2E MP massive MU-MIMO is designed using one deep NN, referred to as AE-MP-mMIMO. Most importantly, the proposed AE-MP-mMIMO is suited for varying block fading channel scenarios, such that it fits any Rayleigh fading and do not need to be retrained when the channel changes. Here, we adopted two-stage precoding scheme, at the relay, to effectively tackle both multiuser interference and nonlinear distortion. It is worth to recall that all the deep NN layers are trained off-line, except the ZF precoder changes on-line when the channel changes. As a perspective, we intend to extend this proposed approach to cell-free massive MIMO systems which can provide uniformly better spectral efficiency for users than traditional cellular technology.

## REFERENCES

- [1] A. Yang, C. Xing, Z. Fei, and J. Kuang, "Performance analysis for uplink massive MIMO systems with a large and random number of UEs," vol. 59, pp. 1–9, 2016.
- [2] F. A. P. de Figueiredo, F. A. C. M. Cardoso, I. Moerman, and G. Fraidenraich, "Channel estimation for massive MIMO TDD systems assuming pilot contamination and frequency selective fading," *IEEE Access*, vol. 5, pp. 17733–17741, 2017.
- [3] P. Pasangi, M. Atashbar, and M. Mohassel Feghhi, "Blind downlink channel estimation for TDD-based multi user massive MIMO in the presence of nonlinear HPA," *Wiley ETRI J.*, vol. 41, pp. 426–436, 2019.
- [4] R. Krishnan, M. R. Khanzadi, N. Krishnan, and Y. Wu, "Linear massive MIMO precoders in the presence of phase noise—A large-scale analysis," *IEEE Trans. Veh. Technol.*, vol. 65, pp. 3057–3071, 2016.
- [5] R. Zayani, H. Shaiek, and D. Roviras, "PAPR-aware massive MIMO-OFDM downlink," *IEEE Access*, vol. 7, pp. 25474–25484, 2019.
- [6] Y. Wang, J. Lei, and F. Shang, "Enabling device-to-device (D2D) communication for the next generation WLAN," *Wiley Hindawi Wireless Commun. Mobile Comput.*, vol. 2021, 2021, Art. no. 1949352.
- [7] H. A. Suraweera, H. Q. Ngo, T. Q. Duong, C. Yuen, and E. G. Larsson, "Multi-pair amplify-and-forward relaying with very large antenna arrays," in *Proc. IEEE Inter. Conf. Commun.*, 2013, pp. 4635–4640.
- [8] H. Cui, L. Song, and B. Jiao, "Multi-pair two-way amplify-and-forward relaying with very large number of relay antennas," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2636–2645, May 2014.
- [9] E. Sharma, A. K. Shukla, and R. Budhiraja, "Spectral- and energy-efficiency of massive MIMO two-way half-duplex hybrid processing AF relay," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 876–879, Oct. 2018.
- [10] J. Li, X. Li, Y. Liu, C. Zhang, L. Li, and A. Nallanathan, "Joint impact of hardware impairments and imperfect channel state information on multi-relay networks," *IEEE Access*, vol. 7, pp. 72358–72375, 2019.
- [11] R. Zayani, R. Bouallegue, and D. Roviras, "Adaptive predistortions based on neural networks associated with Levenberg–Marquardt algorithm for satellite down links," *EURASIP J. Wireless Commun. Netw.*, vol. 2008, no. 1, Jul. 2008, Art. no. 132729.
- [12] J. Tellado, *Multicarrier Modulation With Low PAR: Applications to DSL and Wireless*. New York, NY, USA: Kluwer, 2000.
- [13] H. Li, T. Jiang, and Y. Zhou, "An improved tone reservation scheme with fast convergence for PAPR reduction in OFDM systems," *IEEE Trans. Broadcast.*, vol. 57, no. 4, pp. 902–906, Dec. 2011.
- [14] S. S. K. C. Bulusu, H. Shaiek, D. Roviras, and R. Zayani, "Reduction of PAPR for FBMC-OQAM systems using dispersive SLM technique," in *Proc. 11th Int. Symp. Wireless Commun. Syst.*, 2014, pp. 568–572.
- [15] K. Anoh, C. Tanriover, and B. Adebisi, "On the optimization of iterative clipping and filtering for PAPR reduction in OFDM systems," *IEEE Access*, vol. 5, pp. 12004–12013, 2017.
- [16] S.-H. Wang, W.-L. Lin, B.-R. Huang, and C.-P. Li, "PAPR reduction in OFDM systems using active constellation extension and subcarrier grouping techniques," *IEEE Commun. Lett.*, vol. 20, no. 12, pp. 2378–2381, Dec. 2016.
- [17] S. Shu, D. Qu, L. Li, and T. Jiang, "Invertible subset QC-LDPC codes for PAPR reduction of OFDM signals," *IEEE Trans. Broadcast.*, vol. 61, no. 2, pp. 290–298, Jun. 2015.
- [18] R. Zayani and D. Roviras, "Low-complexity linear precoding for low-PAPR massive MU-MIMO-OFDM downlink systems," *Int. J. Commun. Syst.*, vol. 34, 2021, Art. no. e4889.
- [19] H. Ye, G. Y. Li, and B. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 114–117, Feb. 2018.
- [20] J. C. Shen, J. Zhang, and K. B. Letaief, "Downlink user capacity of massive MIMO under pilot contamination," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3183–3193, Jun. 2015.
- [21] E. Balevi, A. Doshi, and J. G. Andrews, "Massive MIMO channel estimation with an untrained deep neural network," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 2079–2090, Mar. 2020.
- [22] S. Dörner, S. Cammerer, J. Hoydis, and S. ten Brink, "Deep learning based communication over the air," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 132–143, Feb. 2018.
- [23] S. Xue, Y. Ma, and N. Yi, "End-to-end learning for uplink MU-SIMO joint transmitter and non-coherent receiver design in fading channels," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 5531–5542, Sep. 2021.
- [24] J. Vieira, F. Rusek, O. Edfors, S. Malkowsky, L. Liu, and F. Tufvesson, "Reciprocity calibration for massive MIMO: Proposal, modeling, and validation," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 3042–3056, May 2017.
- [25] J. Zhang, X. Xue, E. Björnson, B. Aiand, and S. Jin, "Spectral efficiency of multipair massive MIMO two-way relaying with hardware impairments," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 14–17, Feb. 2018.
- [26] R. Zayani, H. Shaiek, and D. Roviras, "Efficient precoding for massive MIMO downlink under PA nonlinearities," *IEEE Wireless Commun. Lett.*, vol. 23, no. 9, pp. 1453–1456, Sep. 2019.
- [27] J. Bussgang, H. Shaiek, and D. Roviras, "Crosscorrelation functions of amplitude-distorted Gaussian signals," Res. Lab. Electron., Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep. 216, 1952. [Online]. Available: <http://hdl.handle.net/1721.1/4847>
- [28] H. Shaiek, R. Zayani, Y. Medjahdi, and D. Roviras, "Analytical analysis of SER for beyond 5G post-OFDM waveforms in presence of high power amplifiers," *IEEE Access*, vol. 7, pp. 29441–29452, 2019.
- [29] M. Cherif Dakhli, R. Zayani, O. B. Belkacem, and R. Bouallegue, "Theoretical analysis and compensation for the joint effects of HPA nonlinearity and RF crosstalk in VBLAST MIMO OFDM systems over Rayleigh fading channel," *EURASIP J. Wireless Commun.*, vol. 2014, 2014, Art. no. 61.
- [30] H. Lajnef, M. Cherif Dakhli, M. Hizem, and R. Bouallegue, "Compensation schemes and performance analysis of jointly nonlinear amplifier and timing errors for CP-OFDM based cognitive radio networks," *Wireless Networks*, vol. 25, no. 4, pp. 2173–2185, 2019.
- [31] A. Goldsmith, *Wireless Communications*. Stanford, CA, USA: Stanford Univ., 2005.
- [32] F. Sohrabi, H. V. Cheng, and W. Yu, "Robust symbol-level precoding via autoencoder-based deep learning," in *Proc. IEEE Inter. Conf. Acoust. Speech Signal Process.*, Barcelona, Spain, 2020, pp. 8951–8955.
- [33] N. D. Lahbib, M. Cherif, M. Hizem, and R. Bouallegue, "BER analysis and CS-based channel estimation and HPA nonlinearities compensation technique for massive MIMO system," *IEEE Access*, vol. 10, pp. 27899–27911, 2022.