Low-complexity Deep Unfolded Neural Network Receiver for MIMO Systems Based on the Probability Data Association Detector

Pedro Henrique Carneiro Souza (pedro.carneiro@dtel.inatel.br)
Instituto Nacional de Telecomunicacoes https://orcid.org/0000-0002-1905-600X

Luciano Leonel Mendes
Instituto Nacional de Telecomunicacoes

Research Article

Keywords: MIMO, signal detection, machine learning, neural networks, deep unfolding, low-complexity

Posted Date: March 24th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1439479/v1

RESEARCH

Low-complexity Deep Unfolded Neural Network Receiver for MIMO Systems Based on the Probability Data Association Detector

Pedro H. C. de Souza* and Luciano L. Mendes*

*Correspondence:
pedro.carneiro@dtel.inatel.br;
luciano@inatel.br
National Institute of Telecommunications - INATEL, Av. João de Camargo, 510 - Centro, 37540-000 Santa Rita do Sapucaí, BR
Full list of author information is available at the end of the article

Abstract

The interest on applications where machine learning algorithms and communications are combined has been on a rising in recent years. Machine learning and neural networks are being advocated as a way of improving the performance of several functions across all layers of future communication systems. Furthermore, in applications where complexity reduction is essential for the system feasibility at the cost of an affordable performance loss, more efficient systems might be achieved with the aid of machine learning algorithms. Signal detection for multiple-input multiple-output (MIMO) systems has become a hot topic in recent years given its prominent role in fourth and fifth generations of mobile networks. However, the computational complexity in MIMO systems can become prohibitive when the number of antennas is high, as in massive MIMO, for example. Therefore, by leveraging neural networks architectures we propose a deep unfolded detector, whereby the probability data association (PDA) detector algorithm is adapted and enhanced by means of neural network learning capabilities. We unveil that the proposed detector is orders-of-magnitude less complex than the PDA detector and specially than the optimum detector, yet presenting no severe penalties in performance in terms of bit error rate (BER).

Keywords: MIMO; signal detection; machine learning; neural networks; deep unfolding; low-complexity

1 Introduction

The purported success of multiple-input multiple-output (MIMO) systems is being confirmed since the fourth generation of mobile networks (4G) and continue to show

its importance in recent deployments of the fifth generation of mobile networks (5G) technology. Its advantages over classical single-input single-output (SISO) systems are extremely attractive and relatively simple to understand from a theoretical standpoint [1, 2]: by increasing the number of service antennas, an overall increase in data throughput is obtained. More specifically, in the most recent development of multi-antenna systems, known as Massive MIMO [3], dozens of antennas can provide huge gains in performance, simplicity in signal processing and support for a scalable system design.

It was shown in [4] that detectors based on neural networks (NNs) have a competitive performance when compared to the optimum maximum likelihood detector (MLD), while the former is more robust and less complex than the latter. However, the system model in the context of these results considers a SISO system. Recently, several works [5, 6] proposed solutions that attempt to integrate machine learning (ML) and NNs to MIMO systems. One emerging solution involves adapting NN architectures according to model-driven detection algorithms, such that its iterations are unfolded on NN layers. This solution is called deep unfolding.

Therefore, in this work we propose a deep unfolded detector [7] based on the probability data association (PDA) detector [8] for MIMO systems. That way, it is expected that the aforementioned advantages of data-driven detectors in SISO systems could be transferred to MIMO systems, while advantageous features of the PDA detector [2] are maintained.

In this work, the computational complexity of the proposed detector is evaluated and compared with the complexity presented by the MLD and other detectors of interest. Additionally, numerical results resulted from computational simulations, compare the uncoded and coded error rates of the proposed detector with the MLD under time-dispersive channels.

The remainder of this paper is organized as follows. In Section 2, we present the system model of the baseline orthogonal frequency division multiplexing (OFDM)-MIMO system. Section 3 then introduces the problem of signal detection for MIMO systems and gives a brief description of the PDA detector and of the deep unfolding learning. This is followed by a description of the proposed deep unfolded PDA (DU-PDA) and an analysis on the computational complexity of all detectors discussed throughout this paper. Next, in Section 4, we provide numerical results

to evaluate the performance of all detectors studied in this paper, including the optimum MLD. Finally, Section 5 concludes the paper.

1.1 Notation

Throughout this paper, italicized letters (e.g. x or X) represent scalars, boldfaced lowercase letters (e.g. $\mathbf{x}$) represent vectors, and boldfaced uppercase letters (e.g. $\mathbf{X}$) denote matrices. The nth entry of the vector $\mathbf{x}$ is represented by $x(n)$. The entry on the ith row and jth column of the matrix $\mathbf{X}$ is denoted by $X_{i,j}$. The superscript $\mathbf{x}^{(n)}$ denotes the nth instance of the vector $\mathbf{x}$. The sets of real and complex numbers are represented by $\mathbb{R}$ and $\mathbb{C}$, respectively. The absolute value of the scalar $x \in \mathbb{R}$ or the modulo of $x \in \mathbb{C}$ is denoted by $|x|$. The sets of vectors of dimension X with real and complex entries are respectively represented by $\mathbb{R}^X$ and $\mathbb{C}^X$. The sets of matrices of dimension $X \times Y$ with real and complex entries are correspondingly described by $\mathbb{R}^{X\times Y}$ and $\mathbb{C}^{X\times Y}$. The transposition operation of a vector or matrix is represented as $(\cdot)^{\mathrm{T}}$. The ℓ_p-norm, $p \geq 1$, of the vector $\mathbf{x}$ is given by $\|\mathbf{x}\|_p = \left(|x(0)|^p + |x(1)|^p + \cdots + |x(n-1)|^p\right)^{1/p}$. The expected value of the random variable z is denoted by $\mathrm{E}[z]$. The real and imaginary parts of $z \in \mathbb{C}$ are denoted by $\Re(z)$ and $\Im(z)$. The estimate of a scalar x, a vector $\mathbf{x}$ or a matrix $\mathbf{X}$ is represented by $\hat{x}$, $\hat{\mathbf{x}}$ and $\hat{\mathbf{X}}$, respectively. The number of elements in a set $\mathcal{X}$ is given by $\#\mathcal{X}$. Computational complexity is denoted by the asymptotic operator $\mathcal{O}(\cdot)$.

2 System Model

Suppose that in a multiple antenna system we have N_{t} transmitting antennas and N_{r} receiving antennas, thereby constituting a $N_{\mathrm{t}} \times N_{\mathrm{r}}$ MIMO system. Bits of data are demutiplexed into N_{t} substremas, which in turn are mapped to a sequence of complex symbols. These symbols are transmitted by its respective transmit antenna using an OFDM system, for which it is assumed that the cyclic prefix (CP) length is larger than the maximum delay spread for all $N_{\mathrm{t}}N_{\mathrm{r}}$ channels. Finally, after performing the discrete Fourier transform (DFT) we have the following representation of the received baseband signal at the kth subcarrier:

$$\tilde{\mathbf{r}}_k = \tilde{\mathbf{H}}_k \tilde{\mathbf{a}}_k + \tilde{\mathbf{n}}. \tag{1}$$

Here, $\tilde{\mathbf{H}}_k \in \mathbb{C}^{N_\mathrm{r} \times N_\mathrm{t}}$ is the channel frequency response for the kth OFDM subcarrier; $\tilde{\mathbf{a}}_k \in \mathbb{C}^{N_\mathrm{t}}$ represents the symbol vector transmitted by the N_t transmit antennas on the kth subcarrier of the OFDM block and $\tilde{\mathbf{n}} \in \mathbb{C}^{N_\mathrm{r}}$ is the complex additive white Gaussian noise (AWGN) vector for the N_r receive antennas in the frequency domain, with zero mean and covariance matrix given by $\sigma^2 \mathbf{I}_{N_\mathrm{r}}$.

For convenience, henceforth we make use of the real-valued representation [2, 8, 7] for MIMO systems. Therefore, let the received signal (1) be represented by the concatenation of its real and imaginary parts, such that

$$\mathbf{r}_k = \mathbf{H}_k \mathbf{a}_k + \mathbf{n}, \tag{2}$$

where

$$\mathbf{r}_k = \left[\Re(\tilde{\mathbf{r}}_k)^\mathrm{T} \ \Im(\tilde{\mathbf{r}}_k)^\mathrm{T}\right]^\mathrm{T} \in \mathbb{R}^{2N_\mathrm{r}}, \ \forall \ k, \tag{3}$$

$$\mathbf{H}_k = \begin{bmatrix} \Re(\tilde{\mathbf{H}}_k) & -\Im(\tilde{\mathbf{H}}_k) \\ \Im(\tilde{\mathbf{H}}_k) & \Re(\tilde{\mathbf{H}}_k) \end{bmatrix} \in \mathbb{R}^{2N_\mathrm{r} \times 2N_\mathrm{t}}, \ \forall \ k, \tag{4}$$

$$\mathbf{a}_k = \left[\Re(\tilde{\mathbf{a}}_k)^\mathrm{T} \ \Im(\tilde{\mathbf{a}}_k)^\mathrm{T}\right]^\mathrm{T} \in \mathbb{R}^{2N_\mathrm{t}}, \ \forall \ k, \tag{5}$$

$$\mathbf{n} = \left[\Re(\tilde{\mathbf{n}})^\mathrm{T} \ \Im(\tilde{\mathbf{n}})^\mathrm{T}\right]^\mathrm{T} \in \mathbb{R}^{2N_\mathrm{r}}. \tag{6}$$

Moreover, we assume that $\Re(\tilde{\mathbf{a}}_k) \in \mathbb{S}^{N_\mathrm{t}}$ and $\Im(\tilde{\mathbf{a}}_k) \in \mathbb{S}^{N_\mathrm{t}}$, that is, the real and imaginary parts of $\tilde{\mathbf{a}}_k$ can take on different values from the finite set of coordinates pertaining to the square M-quadrature amplitude modulation (QAM) constellation. Hence, let $\mathbb{S} = \{\pm E_0, \pm 3E_0, \ldots, \pm(\sqrt{M}-1)E_0\}$, for $E_0 = \sqrt{\frac{3}{2(M-1)}}$, such that the constellation energy is normalized to 1.

3 Detection in MIMO Systems

A classical problem in the MIMO literature is to decide which symbols were transmitted by each antenna when only possessing (2) at the receiver. This detection problem can be solved optimally, however at great computational effort, by the MLD for MIMO as follows

$$\hat{\mathbf{a}}_k = \underset{\mathbf{a}_k \in \mathbb{S}^{2N_\mathrm{t}}}{\arg\min} \|\mathbf{r}_k - \mathbf{H}_k \mathbf{a}_k\|_2^2, \tag{7}$$

for which $\hat{\mathbf{a}}_k \in \mathbb{R}^{2N_\mathrm{t}}$ is the estimated vector of symbols' coordinates.

It is known that the prohibitive complexity presented by the MLD motivated the research of several alternative detectors for MIMO throughout the last decades [2]. Among them is the PDA detector, whose performance is close to that of the MLD but with a significantly lower complexity, as will be detailed in Subsection 3.5. In the next subsection, the PDA detectors' algorithm is presented, followed by the proposed DU-PDA, for which the PDA is the underlying algorithm.

3.1 Probability Data Association Detector

Before the detection task is carried out by the PDA detector, the received signal, $\mathbf{r}_k$, is preprocessed or equalized using the zero-forcing (ZF) principle as follows [8, 2, 1]

$$\mathbf{z}_k = \mathbf{H}_k^{\dagger}\mathbf{r}_k = \mathbf{a}_k + \mathbf{v}, \tag{8}$$

wherein $\mathbf{H}_k^{\dagger} = (\mathbf{H}_k^T\mathbf{H}_k)^{-1}\mathbf{H}_k^T$ is the left Moore-Penrose psedoinverse and $\mathbf{v} = \mathbf{H}_k^{\dagger}\mathbf{n}$ is the enhanced AWGN. Let us rewrite (8), such that

$$\mathbf{z}_k = \mathbf{e}_i a_k(i) + \underbrace{\sum_{j\neq i} \mathbf{e}_j a_k(j) + \mathbf{v}}_{\mathcal{V}_i},\ \forall\ i, j \in \{0, 1, \ldots, 2N_{\mathrm{t}} - 1\}, \tag{9}$$

where $\mathbf{e}_i$ is the vector with 1 (one) at its ith entry and 0 (zero) otherwise, and $\mathcal{V}_i$ is a multivariate random variable (RV) that can be seen as the effective interference-plus-noise contaminating $a_k(i)$ [8]. Therefore, the crux is at detecting the symbol transmitted by the ith antenna, while considering that all other $j \neq i$ transmitted symbols are interference added to the noise term, which is described by $\mathcal{V}_i$.

Therefore, the PDA detector associates, for each $a_k(i)$, a probability vector $\mathbf{p}_i \in \mathbb{R}^{\sqrt{M}}$, which is given by the evaluation of $P_m(a_k(i) = q\,(m)\ \mid \mathbf{z}_k, \{\mathbf{p}_j\}_{\forall j\neq i})$; $q\,(m) \in \mathbb{S}$ being a coordinate of the M-QAM constellation and $m \in \{0, 1, \ldots, \sqrt{M} - 1\}$. It is important to remark that the PDA detector uses all $\{\mathbf{p}_j\}_{\forall j\neq i}$ associated to interfering symbols already detected, thanks to the incorporation of a strategy similar to that of successive interference cancellation (SIC) detectors. This significantly reduces the computational complexity for calculating $\mathbf{p}_i$, since otherwise $P_m(a_k(i) = q\,(m)\ \mid \mathbf{z}_k)$ would have to be evaluated. The problem here is the requirement of computing multiple integrals for each received symbol, rendering this evaluation prohibitive in practice. Dropping the subscript $(\cdot)_k$ in order to simplify the notation and assuming that $\mathcal{V}_i$ has a Gaussian

distribution [8, 9], then the likelihood function of $\mathbf{z} \mid a(i) = q\,(m)$ can be defined as

$$P_m(\mathbf{z} \mid a(i) = q\,(m)) \propto \exp\left(\alpha_m\,(i)\right), \tag{10}$$

for which,

$$\alpha_m\,(i) = \left(\mathbf{z} - \boldsymbol{\mu}_i - 0.5\mathbf{e}_i q\,(m)\right)^{\mathrm{T}} \boldsymbol{\Omega}_i^{-1}\mathbf{e}_i q\,(m)\,, \tag{11}$$

wherein $\mathrm{E}\left[\mathcal{V}_i\right] = \boldsymbol{\mu}_i$ and $\mathrm{COV}\left[\mathcal{V}_i\right] = \boldsymbol{\Omega}_i$ are given by

$$\boldsymbol{\mu}_i = \sum_{j \neq i} \mathbf{e}_j \left(\mathbf{q}^{\mathrm{T}}\mathbf{p}_j\right), \tag{12}$$

$$\boldsymbol{\Omega}_i = \sum_{j \neq i} \mathbf{e}_j\mathbf{e}_j^{\mathrm{T}} \left(\left(\mathbf{q}^2\right)^{\mathrm{T}}\mathbf{p}_j - \boldsymbol{\mu}_j^2\right) + 0.5\sigma^2\mathbf{G}^{-1}, \tag{13}$$

where $\mathbf{q} = \left[q\,(0)\ \ q\,(1)\ \ \ldots\ \ q(\sqrt{M}-1)\right]^{\mathrm{T}}$ and $\mathbf{G}^{-1} = (\mathbf{H}^{\mathrm{T}}\mathbf{H})^{-1}$ is the inverse of the Gram matrix [1] that accounts for the noise enhancement caused by the ZF. To evaluate the posteriors probabilities associated to each symbol we compute

$$P_m(a(i) = q\,(m)\ \mid \mathbf{z}, \{\mathbf{p}_j\}_{\forall j \neq i}) \approx \frac{P_m(\mathbf{z} \mid a(i) = q\,(m))}{\sum\limits_{m=0}^{\sqrt{M}-1} P_m(\mathbf{z} \mid a(i) = q\,(m))}, \tag{14}$$

which can be seen as an approximate form of the Bayesian theorem [9]. Then substituting (10) into (14) yields

$$p_i\,(m) = \frac{\exp\left(\alpha_m\,(i)\right)}{\sum\limits_{m=0}^{\sqrt{M}-1} \exp\left(\alpha_m\,(i)\right)}. \tag{15}$$

Finally, the PDA detector procedure is given in Algorithm 1.

Note that the optimal detection sequence [8] used in Algorithm 1 can be found with the aid of the following operation:

$$\rho\,(i) = \frac{1}{\mathbf{f}_i^{\mathrm{T}}\mathbf{H}\mathbf{f}_i} \max\left\{0, \mathbf{f}_i^{\mathrm{T}}\mathbf{h}_i - \sum_{j \neq i} |\mathbf{f}_i^{\mathrm{T}}\mathbf{h}_j|\right\}^2, \tag{16}$$

where $\mathbf{f}_i^{\mathrm{T}}$ represents the ith row of $\mathbf{F} = \mathbf{H}^{\dagger}$ and $\mathbf{h}_j$ denotes the jth column of $\mathbf{H}$. Note that larger magnitudes for $\rho\,(i)$ means that the ith antenna suffers less inter antenna interference (IAI) [2]. In other words, the off-diagonal entries of the ith row from $\mathbf{FH}$ have, combined, smaller magnitudes than its

Algorithm 1 The PDA detector

Require: $\tilde{\mathbf{z}}$ via (8)
Require: k_i (see (16)), $\epsilon > 0$
Ensure: $p_i(m) \leftarrow \frac{1}{\sqrt{M}},\ \forall\ m\ \forall\ i$
  repeat
    for $i = 1, 2, \ldots, 2N_t$ **do** ▷ outer iteration
      $\mathbf{p}_i' \leftarrow \mathbf{p}_i$
      Compute $\boldsymbol{\mu}_{k_i}$ and $\boldsymbol{\Omega}_{k_i}$ with $\{\mathbf{p}_j\}_{\forall j \neq k_i}$
      for $m = 1, 2, \ldots, \sqrt{M}$ **do** ▷ inner iteration
        Calculate $\alpha_m(k_i)$ with (11)
        Evaluate:
        $P_m(\tilde{a}(k_i) = q(m) \mid \tilde{\mathbf{z}}, \{\mathbf{p}_j\}_{\forall j \neq k_i}) \approx p_{k_i}(m)$,
        given by (15)
      end for
    end for
  until $|\mathbf{p}_i - \mathbf{p}_i'| \leq \epsilon,\ \forall\ i$ ▷ convergence iteration
  $l_i \leftarrow \arg\max_m \{p_i(m)\},\ \forall\ i$

  Decide transmitted symbols $\hat{a}(i) \leftarrow q_{l_i},\ \forall\ i$

ith diagonal entry. It is easy to show that the optimal sequence is defined by sorting $\boldsymbol{\rho} = [\rho(0)\ \rho(1)\ \ldots\ \rho(2N_\mathrm{t} - 1)]^\mathrm{T}$ in a descending order, denoted as $\{k_i \in \{1, \ldots, 2N_\mathrm{t}\} \mid \rho(k_0) > \rho(k_1) > \ldots > \rho(k_{2N_\mathrm{t}})\}$.

3.2 Deep Unfolding

In general, the NN architecture has shown great potential for detecting signals but its design and parameterization, among other problems, impose limitations [4]. Alternatively, this architecture can be adapted such that iterations of an given algorithm are unfolded on its layers [10, 6, 5], hence the term "unfolding". It is also commonly assumed that the NN employs several layers and, consequently, the term "deep" is added.

More specifically, consider an algorithm with an input vector denoted by $\mathbf{x} \in \mathbb{R}^N$, for which its output is given by $\mathbf{y} \in \mathbb{R}^S$, then this algorithm can be expressed by [10]

$$y(s) = g(\mathbf{x}, \boldsymbol{\psi}, \boldsymbol{\Theta}),\ \forall\ s \in \{0, 1, \ldots, S-1\}, \tag{17}$$

wherein $\boldsymbol{\Theta}$ is the set of all parameters used by the algorithm, $g(\cdot)$ represents a mapping function, usually non-linear, and $\boldsymbol{\psi}$ is iteratively updated as follows

$$\psi_\ell(s) = f(\mathbf{x}, \psi_{\ell-1}(s), \boldsymbol{\Theta}), \tag{18}$$

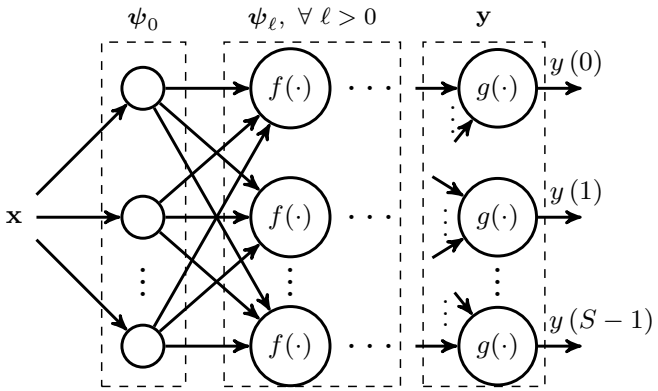

Figure 1 Deep unfolding architecture. It is based on an underlying algorithm with an input vector given by $\mathbf{x}$ and an output determined by $\mathbf{y}$. Each hidden layer unfolds the ℓth iteration of this algorithm and its input-output relationship is expressed by (18), whereas the output layer is represented by (17).

where the ℓth iteration also involves an operation with a mapping function $f(\cdot)$ and $\boldsymbol{\psi}_0$ denotes the initial value.

Therefore, in the deep unfolded context, $\boldsymbol{\psi}_\ell$ can be understood as the input-output relationship at the ℓth layer of a NN architecture, as illustrated in Figure 1. Note that dimensions of learnable parameters $\boldsymbol{\Theta}$ are defined according to the underlying algorithm after which (17), (18), and the architecture depicted in Figure 1 are based. This includes weights and bias, for example, which are optimized by the NN training algorithm [10, 4]. In other words, this means that the number of layers and neurons are fixed, thereby simplifying considerably the process of defining what is commonly known as the NN hyperparemeters.

Moreover, improvements are also obtained by using the aforementioned learnable parameters directly into the iterative algorithm. That way, learning capabilities of NNs can be applied for optimizing algorithms such that its global performance, computational complexity, or even both, are improved. In the next subsection, the PDA detector discussed in Subsection 3.1 is implemented using the deep unfolded architecture for NNs.

3.3 Deep Unfolded PDA Detector

Aiming to take advantage of the iterative algorithm of the PDA detector, we propose the DU-PDA detector. Firstly, in the DU-PDA detector, the received signal, $\mathbf{r}$, is

preprocessed at the ℓth layer by the following operation [7]

$$\mathbf{z}_\ell = \hat{\mathbf{a}}_\ell + w_\ell \mathbf{H}^{\mathrm{T}} \left(\mathbf{r} - \mathbf{H}\hat{\mathbf{a}}_\ell \right), \ \forall \ \ell \in \{0, 1, \ldots, L-1\}, \tag{19}$$

where $\hat{\mathbf{a}}_\ell \in \mathbb{R}^{2N_{\mathrm{t}}}$ is the estimated transmitted symbol vector and the scalar $w_\ell \in \mathbb{R}$ represents a learnable parameter. Note that this preprocessing principle differs from the ZF, which is used by the PDA detector as defined in (8). In contrast, for the proposed DU-PDA, it is employed a preprocessing based on the approximate message passing (AMP) algorithm, which also bear similarities with the Richardson method [1, §IV-6, p. 9]. In this way, $\hat{\mathbf{a}}_\ell$ is updated iteratively until it converges to an acceptable approximation of the transmitted symbol vector. Interestingly, when we have $\hat{\mathbf{a}}_\ell \to \mathbf{a}$, then the so-called residual term $(\mathbf{r} - \mathbf{H}\hat{\mathbf{a}}_\ell) \to \mathbf{n}$, which give us a result in (19) similar to (8).

The preprocessed signal of (19) is then fed into the following operation[1]:

$$\begin{aligned} \psi_{\ell^*}(m) = \operatorname{softm}\Big(& \left(\mathbf{z}_\ell - \boldsymbol{\mu}_{\ell^*} - 0.5\mathbf{e}_{\ell^*} q(m) \right)^{\mathrm{T}} \boldsymbol{\Omega}_{\ell^*}^{-1} \mathbf{e}_{\ell^*} q(m) \Big) \\ & \forall \ m \in \{0, 1, \ldots, \sqrt{M} - 1\}, \end{aligned} \tag{20}$$

where,

$$\operatorname{softm}\left(x_\ell(m) \right) = \frac{e^{x_\ell(m)}}{\sum_{m=0}^{L-1} e^{x_\ell(m)}}. \tag{21}$$

Note that the non-linear function $\operatorname{softm}(\cdot)$ is applied at each layer. This makes (20) identical to (15) except that it is unfolded on successive layers and that $\boldsymbol{\psi}_j = \mathbf{p}_j$. Moreover, since the preprocessing is modified, then it is necessary to redefine the covariance matrix, $\boldsymbol{\Omega}_{\ell^*}$, as follows [11, §III-D, p. 2023], [7]

$$\boldsymbol{\Omega}_{\ell^*} = \sum_{j \neq \ell^*} \mathbf{e}_j \mathbf{e}_j^{\mathrm{T}} \left(\left(\mathbf{q}^2 \right)^{\mathrm{T}} \boldsymbol{\psi}_j - \boldsymbol{\mu}_j^2 \right) + \mathbf{e}_{\ell^*} \mathbf{e}_{\ell^*}^{\mathrm{T}} \operatorname{COV}\left[\mathbf{z}_\ell - \mathbf{a} \right], \tag{22}$$

where,

$$\operatorname{COV}\left[\mathbf{z}_\ell - \mathbf{a} \right] = \frac{[\epsilon_\ell]_+ \| \mathbf{I}_{2N_{\mathrm{t}}} - w_\ell \mathbf{H}^{\mathrm{T}} \mathbf{H} \|_2^2 + 0.5\sigma^2 \| w_\ell \mathbf{H}^{\mathrm{T}} \|_2^2}{2N_{\mathrm{t}}}, \tag{23}$$

wherein $[x]_+ = \max(0, x)$ and for which,

$$\epsilon_\ell = \frac{\| \mathbf{r} - \mathbf{H}\hat{\mathbf{a}}_\ell \|_2^2 - N_{\mathrm{r}}\sigma^2}{\| \mathbf{H} \|_2^2}. \tag{24}$$

[1] $\{\ell^* \in \{0, 1, \ldots, 2N_{\mathrm{t}} - 1\}, \ k \in \{1, 2, \ldots, \lceil L/2N_{\mathrm{t}} \rceil - 1\} \mid \ell^* = \ell - k2N_{\mathrm{t}}; \ k2N_{\mathrm{t}} \leq \ell < (k+1)2N_{\mathrm{t}}\}$

Equation (23) can be understood as the empirical mean-squared error (MSE) estimator of the covariance matrix originated from the residual and noise terms of (19). More importantly, note that $\mathbf{\Omega}_{\ell^*}$ is now a diagonal matrix. This means that computing $\mathbf{\Omega}_{\ell^*}^{-1}$ is not as costly as its counterpart in (11), that is, in the PDA detector. More details about such implications are given in the next subsection.

Therefore, by considering developments presented in this subsection and the general model described in Subsection 3.2, we have

$$\psi_{\ell^*+1}\left(m\right) = \text{softm}\left(\mathbf{z}_\ell, \psi_{\ell^*}\left(m\right), \{w_\ell, \boldsymbol{\mu}_{\ell^*}, \mathbf{\Omega}_{\ell^*}\}\right), \tag{25}$$

which is similar to what is evaluated in (15) with the addition, however, of a learnable parameter and a different preprocessing of the received signal. Note also that $\boldsymbol{\psi}_L = \mathbf{y}$, meaning that the last layer output is also given by (25). Furthermore, let

$$\hat{\mathbf{a}}_{\ell+1} = \sum_{j\neq\ell} \mathbf{e}_j z_\ell\left(j\right) + \mathbf{e}_\ell\left(\mathbf{q}^{\mathrm{T}}\boldsymbol{\psi}_{\ell^*}\right), \tag{26}$$

such that the convergence of (19) might be improved, given that the soft combining of symbols' coordinates and their estimated associated probabilities are fed forward to the next layer.

In Algorithm 2 we detail the general procedure carried out by the proposed DU-PDA detector. The ground truth used for training the NN is defined by $\mathbf{I}_{\ell^*} = [I\left(0\right)\ I\left(1\right)\ \ldots\ I(\sqrt{M}-1)]^{\mathrm{T}}$, such that $\mathbf{I} = \{\mathbf{I}_{\ell^*}\}_{\forall\ell^*}$. It indicates the known constellation coordinates that are transmitted for the training procedure, thus $I_{\ell^*}\left(m\right) \in \{0,1\}\ \forall\ m$. Note also in Algorithm 2 that our formulation of the DU-PDA detector allows usage of the cross-entropy loss function, which contrasts with the popular choice of the MSE loss function [5]. It is a well known fact that the cross-entropy loss function is more appropriate for classification tasks.

3.4 Simplified DU-PDA

The model of the DU-PDA presented in the previous subsection can be simplified even further if some assumptions are made. Therefore, a new variation of the proposed DU-PDA detector, namely the simplified DU-PDA detector, is presented in this subsection. For this detector, the calculations performed in (23) are simplified and the scalar $0.5\sigma^2$ is applied directly in (22). The reasoning behind this approach

Algorithm 2 The DU-PDA detector.

function LAYER($\mathbf{r}$, $\mathbf{H}$, $\boldsymbol{\psi}_{\ell^*-1}$, $\hat{\mathbf{a}}_{\ell-1}$)
  Evaluate (19) and (22), followed by (20) and (26)
  return $\boldsymbol{\psi}_{\ell^*}$, $\hat{\mathbf{a}}_{\ell}$
end function

Ensure: $N_{\mathsf{TR}} > 0$
Ensure: $\psi_{\ell^*}(m) \leftarrow \frac{1}{\sqrt{M}},\ \forall\ m\ \forall\ \ell^*$
Ensure: $\hat{\mathbf{a}}_0 = \sum_{\ell^*} \mathbf{e}_{\ell^*} \left(\mathbf{q}^T \boldsymbol{\psi}_{\ell^*}\right)$
Require: Loss function:
$\mathcal{L}(\mathbf{I}, \boldsymbol{\psi}) = \frac{-1}{\sqrt{M}} \sum_{\ell^*} \mathbf{I}_{\ell^*} \log(\boldsymbol{\psi}_{\ell^*}) + (1 - \mathbf{I}_{\ell^*}) \log(1 - \boldsymbol{\psi}_{\ell^*})$

procedure TRAIN($\mathcal{L}(\mathbf{I}, \boldsymbol{\psi})$, $\boldsymbol{\psi}$, $\hat{\mathbf{a}}_0$, N_{TR})
  for all Epochs **do**
    Generate set of training samples:
    $\mathcal{S}_{\mathsf{TR}} = \left\{\left(\mathbf{r}^{(1)}, \mathbf{I}^{(1)}\right), \ldots, \left(\mathbf{r}^{(N_{\mathsf{TR}})}, \mathbf{I}^{(N_{\mathsf{TR}})}\right)\right\}$
    for $\ell = 1, 2, \ldots, L+1$ **do**
      Train:
      LAYER($\mathbf{r}^{(1,2,\ldots,N_{\mathsf{TR}})}$, $\mathbf{H}^{(1,2,\ldots,N_{\mathsf{TR}})}$, $\boldsymbol{\psi}_{\ell^*-1}$, $\hat{\mathbf{a}}_{\ell-1}$)
    end for
  end for
end procedure

procedure DETECT($\mathbf{r}$, $\mathbf{H}$)
  Execute foward-pass: LAYER($\mathbf{r}$, $\mathbf{H}$, $\boldsymbol{\psi}_{\ell^*-1}$), $\forall\ \ell$
  $d_{\ell^*} \leftarrow \arg\max_m \{\psi_{\ell^*}(m)\},\ \forall\ \ell^*$
end procedure

Decide transmitted symbols $\hat{a}(\ell^*) \leftarrow q_{d_{\ell^*}},\ \forall\ \ell^*$

lies in the asymptotic case, that is, when $N_\mathrm{t} \to \infty$ and $N_\mathrm{r} \to \infty$. For this case, the first term of (23) vanishes, since[2]

$$\mathbf{H}^\mathrm{T}\mathbf{H} \to \mathbf{I}_{2N_\mathrm{t}}, \tag{27}$$

and similarly for the second term we have

$$\|w_\ell \mathbf{H}^\mathrm{T}\|_2^2 \to 2N_\mathrm{t}, \tag{28}$$

which yields

$$\begin{aligned} \mathrm{COV}\left[\mathbf{z}_\ell - \mathbf{a}\right] &\to \frac{[\epsilon_\ell]_+ \|\mathbf{I}_{2N_\mathrm{t}} - \mathbf{I}_{2N_\mathrm{t}}\|_2^2 + N_\mathrm{t}\sigma^2}{2N_\mathrm{t}} \\ &\to 0.5\sigma^2, \end{aligned} \tag{29}$$

wherein, for the sake of simplicity, the learnable parameter w_ℓ is omitted. This is analogous to the channel hardening effect present in massive MIMO systems [2, 1], where values for N_t and N_r are large.

3.5 Computational Complexity

According to the guidelines presented in [4, §IV-C, p. 122404], the global computation complexity of the PDA detector is approximately given by

$$\mathcal{O}(16N_\mathrm{t}^4 + 8\sqrt{M}N_\mathrm{t}^3 + 8N_\mathrm{t}^2(N_\mathrm{r} + \sqrt{M}) + 4N_\mathrm{t}N_\mathrm{r}). \tag{30}$$

However, if we let $N_\mathrm{r} \gg \sqrt{M}$ and simplify constants, then it can be written more compactly as

$$\mathcal{O}(N_\mathrm{t}^4 + \sqrt{M}N_\mathrm{t}^3 + N_\mathrm{t}^2 N_\mathrm{r} + N_\mathrm{t}N_\mathrm{r}). \tag{31}$$

Note that $\mathcal{O}(8N_\mathrm{t}^3 + 16N_\mathrm{t}^2 N_\mathrm{r} + 4N_\mathrm{t}N_\mathrm{r})$ refers to the local cost of (8), where the inverse of $\mathbf{G}$ costs $\mathcal{O}(8N_\mathrm{t}^3)$[3] and $\mathcal{O}(16N_\mathrm{t}^4 + 8\sqrt{M}(N_\mathrm{t}^3 + N_\mathrm{t}^2))$ is the complexity due to computing (11), for which $\mathbf{\Omega}_i^{-1}$ costs $\mathcal{O}(8N_\mathrm{t}^3)$ [8] per outer iteration in Algorithm 1.

[2] We adopt the normalization of the channel matrix by $1/\sqrt{N_\mathrm{r}}$ as is detailed in Section 4.

[3] For the sake of brevity, we assume that the inverse of a matrix, say $\mathbf{X} \in \mathbb{R}^{N\times N}$, is computed by the well-known Gaussian elimination, whose cost is approximately $\mathcal{O}(N^3)$.

Moreover, the DU-PDA detector has an approximate global complexity of

$$\mathcal{O}(4LN_{\mathrm{t}}^2 + 4LN_{\mathrm{t}}(4N_{\mathrm{r}} + \sqrt{M}) + LN_{\mathrm{r}}). \tag{32}$$

Considering again that all constants are simplified and that $N_{\mathrm{r}} \gg \sqrt{M}$, simplifies (32) to

$$\mathcal{O}(LN_{\mathrm{t}}^2 + LN_{\mathrm{t}}N_{\mathrm{r}} + LN_{\mathrm{r}}). \tag{33}$$

The global complexity is composed mainly by the local cost of (19), given by $\mathcal{O}(8N_{\mathrm{t}}N_{\mathrm{r}})$ per layer, and the local cost of (23), expressed by $\mathcal{O}(4N_{\mathrm{t}}^2 + 8N_{\mathrm{t}}N_{\mathrm{r}} + N_{\mathrm{r}})$ for each layer[4]. The NN training stage cost is not taking into account when calculating the computational complexity of the detection stage, since the training stage is assumed to be computed off-line as discussed in [4].

Furthermore, recall that the simplified form of calculation demonstrated by (29) reduces even further the global complexity of the proposed DU-PDA detector. More specifically, the global complexity of the simplified DU-PDA detector is given approximately by $\mathcal{O}(LN_{\mathrm{t}}N_{\mathrm{r}})$, meaning that the cost is reduced to one order-of-magnitude when compared to the DU-PDA detector.

From the computational complexity associated with each detector, it is possible to conclude that the PDA is more complex than the proposed DU-PDA. More specifically, this cost difference is due to the higher order term N_{t}^4, included in the PDA global complexity. This is expected because of the inversion of matrices performed by the PDA detector, which are not necessary for both the DU-PDA and simplified DU-PDA. Also notice that for both of these detectors, the total number of layers L might significantly increase its global complexity. It is demonstrated in the next section, however, that this number is a multiple of N_{t}, thus still implying in a lower global complexity for the DU-PDA when compared to the PDA. In fact, the simplified DU-PDA complexity becomes even lower than that of the ZF in the aforementioned case. Additionally, an optimal detection sequence, such as (16), is not a general requirement for the DU-PDA, which further reduces its global complexity in relation to the PDA.

[4]Note that the squared norm of a matrix $\mathbf{X} \in \mathbb{R}^{M \times N}$ can be written as $\|\mathbf{X}\|_2^2 = \sum_{\forall i} \sum_{\forall j} X_{i,j}^2$, thus its cost is $\mathcal{O}(MN)$.

Table 1 Global computational complexity of detectors studied in this work. Note that they are given in the most compact form and are also ranked in an ascending order, that is, from less to more costly as lines progress to the bottom of the table.

Detector	Global Computational Complexity
Simplified DU-PDA	$\mathcal{O}(LN_{\mathrm{t}}N_{\mathrm{r}})$
Zero Forcing (ZF)	$\mathcal{O}(N_{\mathrm{t}}^3 + N_{\mathrm{t}}^2N_{\mathrm{r}} + N_{\mathrm{t}}N_{\mathrm{r}})$
Deep Unfolded PDA (DU-PDA)	$\mathcal{O}(LN_{\mathrm{t}}^2 + LN_{\mathrm{t}}N_{\mathrm{r}} + LN_{\mathrm{r}})$
Probability Data Association (PDA)	$\mathcal{O}(N_{\mathrm{t}}^4 + \sqrt{M}N_{\mathrm{t}}^3 + N_{\mathrm{t}}^2N_{\mathrm{r}} + N_{\mathrm{t}}N_{\mathrm{r}})$
Maximum Likelihood Detector (MLD)	$\mathcal{O}(M^{N_{\mathrm{t}}}(N_{\mathrm{t}}N_{\mathrm{r}} + N_{\mathrm{r}}))$

For convenience, Table 1 summarizes the global computational complexity for all detectors of interest. To conclude, note also in Table 1 how the complexity of all detectors increase polynomially with the number of transmitting antennas N_{t}. The exception, however, is the MLD, whose complexity increases exponentially with N_{t}, as expected.

4 Numerical Results and Discussion

Before presenting numerical results of detectors performances, we list important system parameters in the following subsection.

4.1 System Parameters

In this work the following system parameters are adopted: (i) before transmission, a frame of n_b data bits is encoded using the polar encoder [12] with a code rate of $R < 1$. Thus, n_b/R bits now represents the coded frame that is effectively transmitted; (ii) entries of the channel frequency response matrix, $\mathbf{H}$, are drawn from a complex Gaussian random process for all k subcarriers at each transmission of an OFDM frame and are normalized by $1/\sqrt{N_{\mathrm{r}}}$. Hence, we have $H_{i,j} \sim \mathcal{CN}\left(0, 1/N_{\mathrm{r}}\right),\ \forall\ i, j$ and, consequently, the system signal-to-noise ratio (SNR) per bit can be expressed as follows

$$\Gamma_k = \left(\sqrt{M}R\right)^{-1} \frac{\mathrm{E}\left[\|\mathbf{H}_k\mathbf{a}_k\|_2^2\right]}{N_{\mathrm{r}}\sigma^2},\ \forall\ k, \tag{34}$$

which is henceforward assumed to be identical for all subcarries.

The bit error rate (BER) is employed for measuring coded detectors' performances, which is obtained by averaging bit decision errors over multiple Monte Carlo experiments. Each experiment is generated using a computational simula-

Table 2 Hyperparamenters of interest for the proposed DU-PDA.

Hyperparameters	Values
Training set size	10^5 samples
Layers	$L = 4N_t$
Input dimension	$\mathbb{R}^{2N_r}$, $\mathbb{R}^{2N_r \times 2N_t}$, $\mathbb{R}^{2N_t \times \sqrt{M}}$, $\mathbb{R}^{2N_t}$
Output dimension	$\mathbb{R}^{2N_t \times \sqrt{M}}$
Number of learnable parameters	$\#\{w_\ell\}_{\forall \ell} = 4N_t$
Activation function	$\mathsf{softmax}(\cdot)$, $\forall \ell$
Learning rate	10^{-3}
Solver	Adam

tion that involves: (i) the generation of $n_b = 256$ equiprobable data bits; (ii) the encoding of data bits by the polar encoder [13]; (iii) mapping of coded bits into complex symbols $\tilde{\mathbf{a}}_k \in \mathbb{S}^{N_\mathrm{t}}$ for all k subcarriers; (iv) transmission of the OFDM frame; (v) the generation of normalized channel coefficients to form entries of the channel matrix $\mathbf{H}_k$; (vi) the generation of complex AWGN samples present in the receiver; (vii) the final decision in favor of the symbol coordinate associated with the higher probability value; (viii) and the subsequent decoding of decided symbols into bits via the polar decoder.

For the sake of brevity, some algorithmic procedures[5] were omitted from Algorithm 2. However, it is worth mentioning that the DU-PDA training is performed considering that SNR values are drawn from a uniform distribution $\mathcal{U} \sim [\min(\mathrm{SNR}), \max(\mathrm{SNR})]$, as discussed in [4, §VI-A, p. 122405]. Additionally, it was decided heuristically to use a total number of $N_\mathrm{TR} = 10^5$ samples for training and also that the DU-PDA should include $L = 4N_\mathrm{t}$ layers[6]. More details about the proposed DU-PDA hyperparameters can be verified in Table 2. These parameters are used for all scenarios demonstrated in the next subsection.

[5] We used the TensorFlow library [14] to implement a customized deep unfolded NN model.
[6] It was verified that the PDA algorithm converges within an average of 2 convergence iterations in Algorithm 1 (with $\epsilon = 10^{-3}$), for all scenarios of interest. Therefore, there is no loss of generality when comparing both detectors costs in the context of results presented in this section.

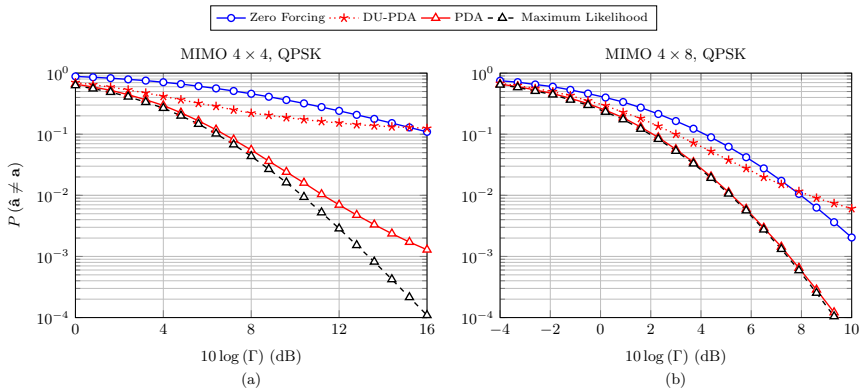

Figure 2 Performance of the ZF, DU-PDA, PDA and MLD detectors for the uncoded $N_\mathrm{t} \times N_\mathrm{r}$ MIMO system. Consequently, the performance metric is the probability of symbol vector error, $P(\hat{\mathbf{a}} \neq \mathbf{a})$, which is given as a function of a range of SNR values. The scenario of (a) 4×4 MIMO is illustrated, followed by the (b) 4×8 MIMO, both considering the QPSK modulation.

4.2 Performance Results

Figure 2 brings the uncoded detection performance for all detectors presented in Table 1, considering a square 4×4 MIMO (Figure 2 (a)) system and a underloaded [2] 4×8 MIMO (Figure 2 (b)), all of which employ the quadrature phase shift keying (QPSK) ($M = 4$) modulation. The detection performance is given as a function of multiple SNR values and it is defined as the probability of occurrence of any error in the received symbol vector. This is done because bits are not encoded for the scenarios analyzed in Figure 2.

Firstly, observe in Figure 2 (a) that the performance of the PDA detector adheres closely with that reported in the seminal work of [8], thus validating the simulation model. Moreover, notice that the DU-PDA detector has shown a prohibitive performance for the 4×4 MIMO scenario, which was also verified to be the case for other square MIMO systems. However, for the underloaded scenario demonstrated in Figure 2 (b), where $N_\mathrm{r} \gg N_\mathrm{t}$, the DU-PDA detector presents better performance. Despite improving its performance relative to other detectors, the proposed DU-PDA detector is still worse than the ZF for SNR > 8 dB. In fact, it was verified that the DU-PDA detector reaches a performance floor of $P(\hat{\mathbf{a}} \neq \mathbf{a}) \approx 3 \times 10^{-3}$, from which no improvement can be obtained irrespective of how high are the SNR values.

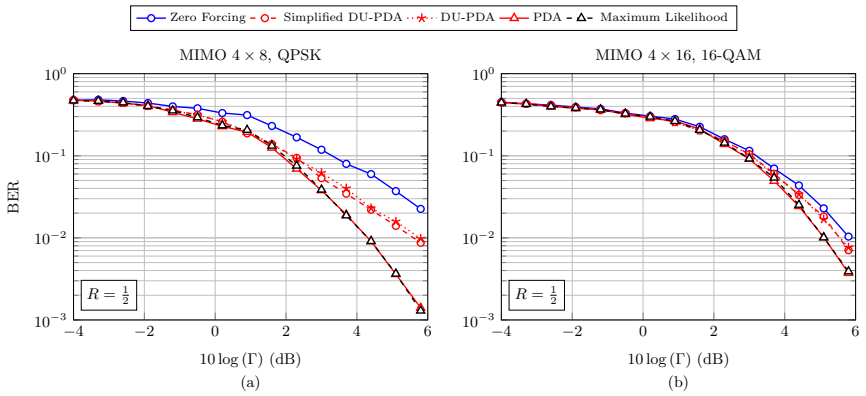

Figure 3 Performance of the ZF, simplified DU-PDA, DU-PDA, PDA and MLD detectors for the coded $N_t \times N_r$ MIMO system. Here, the performance metric is the BER, which is given as a function of a range of SNR values. The scenario presented is of the (a) 4×8 MIMO with a code rate of $R = 1/2$ and QPSK modulation, followed by the (b) 4×16 MIMO also with $R = 1/2$ and considering now the 16-QAM modulation.

This motivated the integration of the Polar encoder as described in Subsection 4.1. Note in Figure 3 (a) that the 4×8 MIMO scenario is illustrated again as in Figure 2, however, considering now the Polar encoding with a code rate of $R = 1/2$. This is accompanied by the Figure 3 (b), for which the 4×16 MIMO scenario with a 16-QAM ($M = 16$) modulation is presented, considering the same aforementioned code rate.

We begin by pointing out that the performance floor observed in Figure 2 (b) is no longer present in Figure 3. More importantly, note in this figure that the detection performance of the DU-PDA is now closer to that of the PDA and Maximum Likelihood detectors. These observations support the conjecture that the uncoded DU-PDA detector is interference limited for high SNR values. In this SNR range the distribution of (19) ceases to be approximately Gaussian because of the low AWGN levels and becomes defined in most part by the non-Gaussian IAI distribution. This in turn violates the Gaussian distribution assumption mentioned in Subsection 3.1, regarding the PDA detector, which is the underlying algorithm of the proposed DU-PDA detector. Hence we have the performance floor shown in Figure 2 (b), but which is successful neutralized with a robust coding scheme.

Moreover, note also that Figure 3 depicts the detection performance of the simplified DU-PDA detector. For this detector, the calculations performed in (23) are simplified, yielding (29). Although the dimensions of MIMO systems illustrated in

Figure 3 are not large, numerical BER results presented here show that conclusions from Subsection 3.4 may still hold for a small number of antennas. Note in Figure 3 that the detection performance of the simplified DU-PDA detector is practically identical to the DU-PDA detectors' performance, for all scenarios analyzed.

Finally, note also that the simplified DU-PDA complexity becomes even lower than that of the ZF, especially when the number of $L = 4N_{\mathrm{t}}$ layers used is considered. This makes the simplified DU-PDA detector the less costly of all detectors analyzed in this work, as can be verified in Table 1, yet it performs approximately 2 dB better than the ZF in Figure 3 (a), for example.

5 Conclusion

In this work we proposed a detector for MIMO systems based upon the deep unfolded architecture for NNs, namely the DU-PDA detector. This detector unfolds iterations of the PDA algorithm in its layers, enhancing the model-driven PDA detector with the aid of its data-driven architecture.

It was shown that the DU-PDA detector has a similar performance, in terms of BER, to that of the optimum detector, that is, the MLD. This can be particularly verified for coded detection in underloaded MIMO systems, for which $N_{\mathrm{r}} \gg N_{\mathrm{t}}$. However, the global computational complexity of the DU-PDA detector is orders-of-magnitude less than the MLD and even the PDA detector, if the simplified DU-PDA is considered. Furthermore, the lack of matrix inverses computations in the DU-PDA detector not only reduces its cost, but also simplifies its implementation in practical systems. This is the case when, for example, channels are correlated, increasing the condition number of $\mathbf{G}$ and making impractical its inverse computation.

Acknowledgements

Not applicable.

Funding

This work was partially supported by RNP, with resources from MCTIC, Grant No. 01245.010604/2020-14, under the 6G Mobile Communications Systems project of the Radiocommunication Reference Center (Centro de Referência em Radiocomunicações - CRR) of the National Institute of Telecommunications (Instituto Nacional de Telecomunicações - Inatel), Brazil and by CNPq-Brazil.

Abbreviations

4G fourth generation of mobile networks

5G fifth generation of mobile networks

AMP approximate message passing

AWGN additive white Gaussian noise

BER bit error rate
CP cyclic prefix
DFT discrete Fourier transform
DU-PDA deep unfolded PDA
IAI inter antenna interference
MIMO multiple-input multiple-output
ML machine learning
MLD maximum likelihood detector
MSE mean-squared error
NN neural network
OFDM orthogonal frequency division multiplexing
PDA probability data association
QAM quadrature amplitude modulation
QPSK quadrature phase shift keying
RV random variable
SIC successive interference cancellation
SISO single-input single-output
SNR signal-to-noise ratio
ZF zero-forcing

Authors' contributions

Both authors contributed equally for this publication.

Authors' information

P. H. C. S. was born in Santa Rita do Sapucaí, Minas Gerais, MG, Brazil in 1992. He received the B.S. and M.S. degrees in telecommunications engineering from the National Institute of Telecommunications - INATEL, Santa Rita do Sapucaí, in 2015 and 2017, respectively; is currently working toward the PhD degree in telecommunications engineering at INATEL.

During the year of 2014 he was a Hardware Tester with the INATEL Competence Center - ICC. His main interests are: digital communication systems, mobile telecommunications systems, 6G, cognitive radio, convex optimization for telecommunication systems, compressive sensing/learning, embedded systems and embedded hardware/firmware.

L. L. M. received the B.Sc. and M.Sc. degrees from Inatel, Brazil, in 2001 and 2003, respectively, and the Doctor degree from Unicamp, Brazil, in 2007, all in electrical engineering. Since 2001, he has been a Professor with Inatel, where he has acted as the Technical Manager of the Hardware Development Laboratory from 2006 to 2012. From 2013 to 2015, he was a Visiting Researcher with the Technical University of Dresden in the Vodafone Chair Mobile Communications Systems, where he has developed his postdoctoral. In 2017, he was elected Research Coordinator of the 5G Brazil Project, an association involving industries, telecom operators, and academia which aims for funding and build an ecosystem toward 5G in Brazil. He is also the technical coordinator of the Brazil 6G Project.

Author details

National Institute of Telecommunications - INATEL, Av. João de Camargo, 510 - Centro, 37540-000 Santa Rita do Sapucaí, BR.

References

1. Albreem, M.A., Juntti, M., Shahabuddin, S.: Massive MIMO detection techniques: A survey. IEEE Communications Surveys & Tutorials **21**(4), 3109–3132 (2019). doi:10.1109/COMST.2019.2935810
2. Yang, S., Hanzo, L.: Fifty years of MIMO detection: The road to large-scale MIMOs. IEEE Communications Surveys & Tutorials **17**(4), 1941–1988 (2015). doi:10.1109/COMST.2015.2475242
3. Marzetta, T.L.: Massive MIMO: An introduction. Bell Labs Technical Journal **20**, 11–22 (2015). doi:10.15325/BLTJ.2015.2407793

4. De Souza, P.H.C., Mendes, L.L., Chafii, M.: Compressive learning in communication systems: A neural network receiver for detecting compressed signals in OFDM systems. IEEE Access **9**, 122397–122411 (2021). doi:10.1109/ACCESS.2021.3108061
5. Balatsoukas-Stimming, A., Studer, C.: Deep unfolding for communications systems: A survey and some new directions. In: 2019 IEEE International Workshop on Signal Processing Systems (SiPS), pp. 266–271 (2019). doi:10.1109/SiPS47522.2019.9020494
6. Pham, Q.-V., Nguyen, N.T., Huynh-The, T., Le, L.B., Lee, K., Hwang, W.-J.: Intelligent Radio Signal Processing: A Contemporary Survey. arXiv:2008.08264 (2020). 2008.08264
7. Khani, M., Alizadeh, M., Hoydis, J., Fleming, P.: Adaptive neural signal detection for massive MIMO. IEEE Transactions on Wireless Communications **19**(8), 5635–5648 (2020). doi:10.1109/TWC.2020.2996144
8. Pham, D., Pattipati, K.R., Willett, P.K., Luo, J.: A generalized probabilistic data association detector for multiple antenna systems. In: 2004 IEEE International Conference on Communications (IEEE Cat. No.04CH37577), vol. 6, pp. 3519–35226 (2004). doi:10.1109/ICC.2004.1313198
9. Yang, S., Lv, T., Maunder, R.G., Hanzo, L.: From nominal to true a posteriori probabilities: An exact bayesian theorem based probabilistic data association approach for iterative MIMO detection and decoding. IEEE Transactions on Communications **61**(7), 2782–2793 (2013). doi:10.1109/TCOMM.2013.053013.120427
10. Zappone, A., Renzo, M.D., Debbah, M.: Wireless Networks Design in the Era of Deep Learning: Model-Based, AI-Based, or Both? arXiv:1902.02647 (2019). 1902.02647
11. Ma, J., Ping, L.: Orthogonal AMP. IEEE Access **5**, 2020–2033 (2017). doi:10.1109/ACCESS.2017.2653119
12. Guimarães, D.A.: Digital Transmission: A Simulation-Aided Introduction with VisSim/Comm. Springer, Berlin Heidelberg (2009). doi:10.1007/978-3-642-01359-1
13. Besser, K.: Digcommpy 0.8. `https://pypi.org/project/digcommpy/` Accessed 2021-09-21
14. et al, M.A.: TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. Software available from tensorflow.org (2015). `https://www.tensorflow.org/`