Symbol Detection Performance and Complexity in Large-scale MIMO Systems

Pedro H. C. Souza, Luciano Mendes, Davi Brilhante and José Rezende

***Abstract*— This paper presents an unified analysis of multiple-input multiple-output (MIMO) detectors which aims to shed light on the compromise between complexity and symbol error rate (SER) performance, showing the conditions in which each detector is more interesting. To demonstrate this unified approach, five detectors with different levels of complexity and performance are evaluated under a large-scale MIMO scenario.**

***Keywords*— MIMO-OFDM detectors, computational complexity, symbol error rate, large-scale MIMO.**

I. Introduction

The high demand for data rates in wireless communication networks in the last decades has pushed the development of multiple-input multiple-output (MIMO) systems for mobile networks. The advent of fifth generation of mobile network (5G) systems has introduced different operational modes where very high throughput allied with very low latency is supported, all this considering the massive connectivity of devices to the base station (BS) [1]. Besides, the recent conception of sixth generation of mobile network (6G) is also pushing the number of transmitting and receiving antennas in MIMO system, leading to the so called large-scale MIMO or massive MIMO [2].

The maximum likelihood detector (MLD) is the benchmark detector for MIMO systems, since it can achieve the best symbol error rate (SER) performance by harvesting all diversity available in the communication channel. However, it is prohibitively complex, thus hindering its use in practical systems. On the other hand, linear detectors based on the channel matrix inversion, such as zero-forcing (ZF) and minimum mean square error (MMSE), have low complexity, however at the cost of poor SER performance, since these detectors are not able to harvest any diversity from the channel. Therefore, the sphere detector (SD) has been proposed as a way to achieve the MLD performance with reduced complexity, by restraining the search for the most likely transmitted sequence to a set within the hypersphere [3]. However, the fixed-complexity SD [4], for example, presents a much smaller complexity than the MLD, but still has a complexity that grows exponentially with the square root of the number of transmit antennas. Therefore, many MIMO detectors have been proposed in order to achieve a better trade-off between performance and complexity [4]. Among these, the approximate message-passing (AMP) [5] excels for its very low complexity, while the probability data association (PDA) [6] achieves acceptable performance with a polynomial complexity.

In this work, the SER performance and the respective computational complexity of baseband MIMO detectors are studied, assuming a MIMO system that scales from 4×4 up to 32×64 antennas. Moreover, a rich scattering channel environment is considered, which motivated the use of orthogonal frequency division multiplexing (OFDM) [7] for symbol transmission. Consequently, a flat Rayleigh channel can be assumed for each OFDM subcarrier without loss of generality, considering also that the channel is constant over the duration of an OFDM frame. In this context, five detectors are compared in this paper, namely, MLD as a benchmark, SD, AMP, PDA, and ZF. The main contribution of this work is to present an unified analysis that takes into account both SER performance and complexity metrics and, consequently, to compare these different detectors under the same channel conditions. This approach allows for an analysis of the trade-off between complexity and SER performance, showing the conditions in which each detector is more interesting.

In order to achieve this goal, this paper is organized as follows: Section II brings the system model used to draw a fair comparison of all detectors, Section III describes the principles of each detector, Section IV brings the complexity and SER performance analysis, and Section V concludes the paper.

II. System Model

We assume a $N_\mathrm{t} \times N_\mathrm{r}$ ($N_\mathrm{t} \leq N_\mathrm{r}$) point-to-point baseband MIMO system, where N_t denotes the number of transmitting antennas and N_r the number of receiving antennas, respectively. Therefore, a bit stream is demultiplexed into N_t substreams, which in turn are mapped into a sequence of complex M-quadrature amplitude modulation (QAM) symbols. These symbols are transmitted by their respective transmitting antenna using OFDM, 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.

Consequently, for a given OFDM symbol, N_t M-QAM symbols are transmitted in the same subcarrier. Hence, after the discrete Fourier transform (DFT) operation on the receiver, the received baseband signal at the kth subcarrier is given by

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

for which $\tilde{\mathbf{H}}_k \in \mathbb{C}^{N_\mathrm{r} \times N_\mathrm{t}}$ is the matrix containing all channel frequency responses for the kth OFDM subcarrier; $\tilde{\mathbf{a}}_k \in \mathbb{C}^{N_\mathrm{t}}$ represents the symbol vector transmitted by the N_t transmitting antennas on the kth subcarrier and $\tilde{\mathbf{n}}_k \in \mathbb{C}^{N_\mathrm{r}}$ is the complex additive white Gaussian noise (AWGN) vector at the kth

Pedro H. C. Souza and Luciano Mendes, Inatel, Sta. Rita do Sapucaí-MG, e-mails: {pedro.carneiro@dtel and luciano@}.inatel.br; Davi Brilhante and José Rezende, UFRJ, Rio de Janeiro-RJ, e-mails: {dbrilhante and rezende}@land.ufrj.br. This work was partially funded by Brasil 6G project (RNP/MCTI grant 01245.010604/2020-14), SAMURAI project (FAPESP grant 20/05127-2), FAPEMIG and CNPq-Brasil.

subcarrier for the N_r receiving antennas, with zero mean and covariance matrix given by $\sigma^2\mathbf{I}_{N_\mathrm{r}}$. For convenience, consider the real-valued representation [4] for MIMO systems, which allows us to rewrite (1) as

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

where

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

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

with $\mathbf{a}_k$ and $\mathbf{n}_k$ having the same structure as $\mathbf{r}_k$, but with the real and imaginary values of $\tilde{\mathbf{a}}_k$ and $\tilde{\mathbf{n}}_k$, respectively. 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-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 (unity).

III. MIMO Detectors

A. Maximum Likelihood

The MLD estimates the symbol vector by solving: $\arg\min \|\mathbf{r}_k - \mathbf{H}_k\mathbf{a}_k\|_2^2$, for which $\mathbf{a}_k, \hat{\mathbf{a}}_k \in \mathbb{S}^{2N_\mathrm{t}}$. Although achieving the best SER performance, the associated high complexity hinders its feasibility in practical systems. Next, sub-optimal detectors with more affordable complexity are described.

B. Sphere Detector

The SD is based on a space reduced search over a N_t-dimensional hypersphere with radius d. Thus, the SD considers only the solutions for

$$\|\mathbf{r}_k - \mathbf{H}_k\mathbf{a}_k\|^2 \leq d^2. \tag{5}$$

However, determining the ideal search radius and the points inside a N_t-dimensional sphere can still be a complex task. To reduce the complexity, a wise approach is to iteratively increment the dimension of the hypersphere, as the points inside of a $k-1$ dimensional sphere are still inside of a k dimensional sphere and as lower the dimension gets the less complex the search becomes [3]. We can use a $\mathbf{QR}$ decomposition on $\mathbf{H}_k$ to exploit some useful properties. Firstly, $\mathbf{R}^{N_\mathrm{t}\times N_\mathrm{t}}$ is upper right triangular; Secondly, $\mathbf{Q}^{N_\mathrm{r}\times N_\mathrm{r}}$ is an orthonormal matrix. We also introduce $\hat{\mathbf{a}}_k = \mathbf{H}_k^\dagger\mathbf{r}_k$, the least-squares estimation of the transmitted symbol $\mathbf{a}_k$, where the Moore-Penrose pseudo inverse is defined by $\mathbf{H}_k^\dagger = \left(\mathbf{H}_k^H\mathbf{H}_k\right)^{-1}\mathbf{H}_k$, and apply in (5), leading to

$$d^2 - \|\mathbf{r}_k\|^2 + \|\mathbf{H}_k\hat{\mathbf{a}}_k\|^2 \geq \|\mathbf{R}(\hat{\mathbf{a}}_k - \mathbf{a}_k)\|^2. \tag{6}$$

The upper triangular $\mathbf{R}$ induces a tree with N_t levels and M branches per node. Thus, the symbol estimation based on (6) can be performed as a depth-first search in this tree [4] and

$$d'^2 \geq \sum_{i=1}^{N_\mathrm{t}}\sum_{j=i}^{N_\mathrm{t}} R_{i,j}^2[(\hat{a}_k)_j - (a_k)_j]^2, \tag{7}$$

where, $d'^2 = d^2 - \|\mathbf{r}_k\|^2 + \|\mathbf{H}_k\hat{\mathbf{a}}_k\|^2$. Taking the square root on both sides and expanding the right-hand side, we have a lower bound and an upper bound for $(a_k)_j$, whereby equation (8) is a recurrence formula for the lower and upper bounds:

$$\left\lceil (\hat{a}_k)_{i|i+1} - \frac{d'_i}{R_{i,i}} \right\rceil \leq (a_k)_i \leq \left\lfloor (\hat{a}_k)_{i|i+1} + \frac{d'_i}{R_{i,i}} \right\rfloor, \tag{8}$$

wherein

$$(\hat{a}_k)_{i|i+1} = (\hat{a}_k)_i - \sum_{j=i+1}^{N_\mathrm{t}} \frac{R_{i,j}}{R_{i,i}}((a_k)_j - (\hat{a}_k)_j), \tag{9}$$

and

$$d'^2_i = d'^2_{i+1} - R^2_{i+1,i+1}[(a_k)_{i+1} - (\hat{a}_k)_{i+1|i+2}]^2. \tag{10}$$

Once the bounds are evaluated, the candidates for each $(a_k)_i$ are the constellation symbols that lies in between the upper and lower bounds. When all the symbols have a candidate, that is, the detector reached $(a_k)_1$, the search process starts over again with a new radius equal to the final distance, $d'^2_{N_\mathrm{t}} - d'^2_1 + R^2_{1,1}[(a_k)_1 - (\hat{a}_k)_{1|2}]^2$, reducing the tree length. If there is no candidate for some $(a_k)_i$, the detector assesses the previous candidates for $(a_k)_{i+1}$, $(a_k)_{i+2}$ and so on. The algorithm stops when the search backtracks to $(a_k)_{N_\mathrm{t}}$ and there is no other possible candidate solution, as described in Algorithm 1.

Algorithm 1 The Sphere Detector

Require: $\mathbf{r}_k$, d, $\mathbf{QR} = \mathbf{H}_k$, $\hat{\mathbf{a}}_k = \mathbf{H}_k^\dagger\mathbf{r}_k$
1: $i \leftarrow N_\mathrm{t}$, $(\hat{a}_k)_{i|i+1} = (\hat{a}_k)_{N_\mathrm{t}}$
2: $d'^2 \leftarrow d^2 - \|\mathbf{r}_k\|^2 + \|\mathbf{H}_k\hat{\mathbf{a}}_k\|^2$
3: Compute $LB[(a_k)_{N_\mathrm{t}}]$ and $UB[(a_k)_{N_\mathrm{t}}]$ from (8)
4: $(a_k)_i \leftarrow LB[(a_k)_i] - 1$
5: **repeat**
6: $\quad(a_k)_i \leftarrow (a_k)_i + 1$
7: $\quad$**if** $(a_k)_i \leq UB[(a_k)_i]$ **then**
8: $\quad\quad$**if** $i > 1$ **then**
9: $\quad\quad\quad i \leftarrow i - 1$
10: $\quad\quad\quad$Compute $(\hat{a}_k)_{i|i+1}$ from (9) and d'^2_i from (10)
11: $\quad\quad\quad$Compute $LB[(a_k)_i]$ and $UB[(a_k)_i]$ from (8)
12: $\quad\quad\quad(a_k)_i \leftarrow LB[(a_k)_i] - 1$
13: $\quad\quad$**else**
14: $\quad\quad\quad$Save $\mathbf{a}_k$, $d'^2_{N_\mathrm{t}} - d'^2_1 + R^2_{1,1}[(a_k)_1 - (\hat{a}_k)_{1|2}]^2$
15: $\quad\quad$**end if**
16: $\quad$**else**
17: $\quad\quad i \leftarrow i + 1$
18: $\quad$**end if**
19: **until** $i = N_\mathrm{t} + 1$

The SD performance is tightly coupled to the choice of the initial radius. If the radius is too large, the search will not be efficient. On the other hand, if the radius is too short, then there will be no candidates within the sphere.

C. Zero Forcing

The ZF is based on equalizing (2) by the inverse of the channel matrix. If $\mathbf{H}_k$ is not square, the Moore-Penrose pseudo

inverse, $\mathbf{H}_k^\dagger$, can be used to achieve the equalizing matrix. Then, the estimated symbol is equal to the constellation symbol with the shorter distance to the equalized received symbol:

$$\hat{\mathbf{a}}_k = \underset{\mathbf{a}_k \in \mathbb{S}^{2N_\mathrm{t}}}{\arg\min} \|\mathbf{H}_k^\dagger \mathbf{r}_k - \mathbf{a}_k\|_2^2, \tag{11}$$

also known as the slicing operation $\mathcal{S}(\mathbf{H}_k^\dagger \mathbf{r}_k)$. Applying (2) in (11) leads to

$$\mathbf{H}_k^\dagger \mathbf{r}_k - \mathbf{a}_k = (\mathbf{I}_{N_\mathrm{t}} \mathbf{a}_k + \mathbf{H}_k^\dagger \mathbf{n}_k) - \mathbf{a}_k, \tag{12}$$

where we can notice that the ZF detector nullifies the effect of interfering symbols, as the identity matrix is expected as the result of the product between the pseudo-inverse and the channel matrix [4]. On the other hand, the noise can be amplified if $\det\left(\mathbf{H}_k^T \mathbf{H}_k\right)$ is small, which is the case when the channel presents severe fading. Consequently, the noise enhancement can severely affect the ZF SER performance in time-variant non-line-of-sight (NLOS) channels.

Algorithm 2 describes ZF detection procedure.

Algorithm 2 The Zero-Forcing Detector

Require: $\mathbf{H}_k, \mathbf{r}_k$
1: $\mathbf{H}_k^\dagger \leftarrow (\mathbf{H}_k^H \mathbf{H}_k)^{-1} \mathbf{H}_k^H$
2: $\hat{\mathbf{r}}_k \leftarrow \mathbf{H}_k^\dagger \mathbf{r}_k$
3: **for** $\hat{r}_k =$ each line in $\hat{\mathbf{r}}_k$ **do**
4: $\quad \hat{a}_k = \underset{\mathbf{a}_k \in \mathbb{S}^{2N_\mathrm{t}}}{\arg\min} |\hat{r}_j - \mathbf{a}_k|^2$
5: **end for**

D. Probability Data Association

Before the detection task is carried out by the PDA detector, the received signal, $\mathbf{r}_k$, is equalized by the ZF, leading to

$$\mathbf{z}_k = \mathbf{H}_k^\dagger \mathbf{r}_k = \mathbf{a}_k + \mathbf{v}, \tag{13}$$

where $\mathbf{v} = \mathbf{H}_k^\dagger \mathbf{n}$ is the enhanced AWGN.

Rewriting (13), leads to

$$\mathbf{z}_k = \mathbf{e}_i a_k(i) + \underbrace{\sum_{j \neq i} \mathbf{e}_j a_k(j) + \mathbf{v}}_{\mathcal{V}_i}, \ \forall\, i, j \in \{0, 1, \ldots, 2N_\mathrm{t} - 1\}, \tag{14}$$

where $\mathbf{e}_i$ is the vector with 1 (one) at its ith entry and 0 (zero) otherwise and $\mathcal{V}_i$ is the effective noise contaminating $a_k(i)$ [6]. The main challenge is to detect the symbol transmitted by the ith antenna, while considering that all other $j \neq i$ transmitted symbols are interference added to $\mathcal{V}_i$.

The PDA detector associates a probability vector $\mathbf{p}_i \in \mathbb{R}^{\sqrt{M}}$ for each $a_k(i)$, 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 use all $\{\mathbf{p}_j\}_{\forall j \neq i}$ associated to interfering symbols already detected, thanks to its incorporation of a strategy similar to that of successive interference cancellation (SIC) detectors. For the sake of simplicity, the subscript $(\cdot)_k$ denoting the kth subcarrier will be omitted at this point.

Assuming that $\mathcal{V}_i$ has a Gaussian distribution [6], [8], then the likelihood function of $\mathbf{z} \mid a(i) = q_m$ can be defined as

$$P_m(\mathbf{z} \mid a(i) = q_m) \propto \exp\left(\alpha_m\left(i\right)\right), \tag{15}$$

for which,

$$\alpha_m\left(i\right) = \left(\mathbf{z} - \boldsymbol{\mu}_i - 0.5\mathbf{e}_i q_m\right)^T \boldsymbol{\Omega}_i^{-1} \mathbf{e}_i q_m, \tag{16}$$

wherein

$$\boldsymbol{\mu}_i = \sum\nolimits_{j \neq i} \mathbf{e}_j \left(\mathbf{q}^T \mathbf{p}_j\right), \text{where } \mathbf{q} = [q_0 \ q_1 \ \ldots \ q_{\sqrt{M}-1}]^T; \tag{17}$$

$$\boldsymbol{\Omega}_i = \sum\nolimits_{j \neq i} \mathbf{e}_j \mathbf{e}_j^T \left(\left[\mathbf{q}^2\right]^T \mathbf{p}_j - \boldsymbol{\mu}_j^2\right) + 0.5\sigma^2 \mathbf{G}^{-1}, \tag{18}$$

and $\mathbf{G}^{-1} = (\mathbf{H}^T \mathbf{H})^{-1}$ is the inverse of the Gram matrix [9] that accounts for the ZF noise enhancement. The posteriors probabilities associated to each symbol is given by [8]

$$p_i\left(m\right) = \frac{\exp\left(\alpha_m\left(i\right)\right)}{\sum_{m=0}^{\sqrt{M}-1} \exp\left(\alpha_m\left(i\right)\right)}. \tag{19}$$

Algorithm 3 describes the procedure for the PDA detection.

Algorithm 3 The PDA detector

Require: $\tilde{\mathbf{z}}$ via (13)
Require: k_i (refer to [10, §II-C, p. 222]), $\epsilon > 0$
Ensure: $p_i\left(m\right) \leftarrow \frac{1}{\sqrt{M}}, \ \forall\, m \ \forall\, i$
1: **repeat**
2: $\quad$ **for** $i = 1, 2, \ldots, 2N_\mathrm{t}$ **do** ▷ outer iteration
3: $\quad\quad \mathbf{p}_i' \leftarrow \mathbf{p}_i$
4: $\quad\quad$ Compute $\boldsymbol{\mu}_{k_i}$ from (17)
5: $\quad\quad$ Compute $\boldsymbol{\Omega}_{k_i}$ from (18)
6: $\quad\quad$ **for** $m = 1, 2, \ldots, \sqrt{M}$ **do** ▷ inner iteration
7: $\quad\quad\quad$ Calculate $\alpha_m\left(k_i\right)$ from (16)
8: $\quad\quad\quad$ Evaluate (19):
9: $\quad\quad\quad$ $P_m(a(k_i) = q\left(m\right) \mid \mathbf{z}, \{\mathbf{p}_j\}_{\forall j \neq k_i}) \approx p_{k_i}\left(m\right)$
10: $\quad\quad$ **end for**
11: $\quad$ **end for**
12: **until** $|\mathbf{p}_i - \mathbf{p}_i'| \leq \epsilon, \ \forall\, i$ ▷ convergence iteration
13: $l_i \leftarrow \underset{m}{\arg\max}\{p_i\left(m\right)\}, \ \forall\, i$
14: Decide transmitted symbols $\hat{a}(i) \leftarrow q_{l_i}, \ \forall\, i$

E. Approximate Message-Passing

The AMP detector is an approximation of the seminal messaging passing algorithm, largely used in the context of graph representations [11]. The AMP detector consists of the following sequence of updates:

$$\begin{aligned} \mathbf{z}_t &= \hat{\mathbf{a}}_t + \mathbf{H}^T\left(\mathbf{r} - \mathbf{H}\hat{\mathbf{a}}_t\right) + \mathbf{b}_t; \\ \mathbf{b}_t &= \frac{\tau_t}{1 + \tau_{t-1}}\left[\mathbf{H}^T\left(\mathbf{r} - \mathbf{H}\hat{\mathbf{a}}_{t-1}\right) + \mathbf{b}_{t-1}\right]; \\ \hat{\mathbf{a}}_{t+1} &= \eta_t\left(\mathbf{z}_t \ ; \ \sigma_t\right); \end{aligned} \tag{20}$$

where $\mathcal{S}\left(\hat{\mathbf{a}}_t\right) \in \mathbb{S}^{2N_\mathrm{t}}$ is the estimated transmitted symbol vector and $\sigma_t = 0.5\sigma^2\left(1 + \tau_t\right)$. Additionally, we have

$$\tau_t = \frac{2N_\mathrm{t}}{\sigma^2 N_\mathrm{r}} \left\langle \eta_t' \right\rangle, \tag{21}$$

wherein $\langle v \rangle = N^{-1} \sum_{i=0}^{N-1} v\,(i)$ and

$$\eta_t' = \sum_{q_m \in \mathbb{S}} q_m^2 P_m(\mathbf{a}_t = q_m \mid \mathbf{z}_t) - \left[\overbrace{\sum_{q_m \in \mathbb{S}} q_m P_m(\mathbf{a}_t = q_m \mid \mathbf{z}_t)}^{\eta_t} \right]^2 . \tag{22}$$

To elaborate, η_t is the optimal Gaussian denoising function, where $f(x)$ is operated element-wise for the vector obtained and thus $f : \mathbb{R} \to \mathbb{R}$. Therefore, it is a non-linear operation that improves the quality of the estimates, $\hat{\mathbf{a}}_t$, across multiple iterations of (20). Interestingly, when we have $\hat{\mathbf{a}}_t \to \mathbf{a}$, then the residual term $(\mathbf{r} - \mathbf{H}\hat{\mathbf{a}}_t) \to \mathbf{n}$ and $\mathbf{b}_t \to 0$, which give us a result in (20) similar to (11).

The main feature of the AMP detector is the lack of matrix inverses computations. Consequently, not only the global computational complexity is reduced, but also the implementation in practical systems is simplified. The AMP detector is an interesting solution when the channel matrix is ill conditioned and its inverse cannot be obtained.

IV. Numerical Results

In this work, entries of $\mathbf{H}$ 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 $\tilde{H}_{i,j} \sim \mathcal{CN}\,(0, 1/N_\mathrm{r})\,,\ \forall\ i, j$ and, consequently, the signal-to-noise ratio (SNR) per bit can be expressed as follows,

$$\Gamma_k = \frac{\mathrm{E}\left[\|\mathbf{H}_k \mathbf{a}_k\|_2^2\right]}{N_\mathrm{r}\sqrt{M}\sigma^2},\ \forall\ k, \tag{23}$$

which is assumed to be identical for all subcarries. Note, however, that the instantaneous channel gains differ among subcarriers at a given symbol transmission cycle. The probability of symbol vector error, $P\,(\hat{\mathbf{a}} \neq \mathbf{a})$, obtained by averaging the results from multiple Monte Carlo simulations, represents the performance metric of the detectors. Moreover, the total number of Monte Carlo trials was attached to the symbol vector errors, such that new trials were interrupted as the mark of 10^4 errors was reached.

A. Computational Complexity

The MLD computational complexity is given by $\mathcal{O}(M^{N_\mathrm{t}})$ [4], whereas for the SD we considered $\mathcal{O}(M^{\sqrt{N_\mathrm{t}}})$ from the fixed-complexity SD [4, §VIII-D, p. 20], since its performance is near–optimum. Furthermore, according to the guidelines presented in [12, §IV-C, p. 122404], the computation complexity of the PDA detector is approximately given by

$$\mathcal{O}(N_\mathrm{t}^4 + \sqrt{M}N_\mathrm{t}^3 + N_\mathrm{t}^2 N_\mathrm{r} + N_\mathrm{t} N_\mathrm{r}), \tag{24}$$

per convergence iteration. Note that

$$\mathcal{O}(8N_\mathrm{t}^3 + 16N_\mathrm{t}^2 N_\mathrm{r} + 4N_\mathrm{t} N_\mathrm{r}), \tag{25}$$

refers to the local cost of (13) or, equivalently, of the ZF detector given by (11). Moreover, the inverse of $\mathbf{G}$ costs $\mathcal{O}(8N_\mathrm{t}^3)$ and $\mathcal{O}(16N_\mathrm{t}^4 + 8\sqrt{M}(N_\mathrm{t}^3 + N_\mathrm{t}^2))$ is the complexity due to computing (16), for which $\mathbf{\Omega}_i^{-1}$ costs $\mathcal{O}(8N_\mathrm{t}^3)$ [6] per outer iteration in Algorithm 3. It was verified that the PDA algorithm converges within an average of less than 2 iterations in Algorithm 3, considering $\epsilon = 10^{-3}$ and $P\,(\hat{\mathbf{a}} \neq \mathbf{a}) < 10^{-2}$, for all scenarios of interest. Hence, we assume the PDA complexity per iteration is approximately the global PDA complexity. Finally, for the AMP detector, N_I refers to the number of iterations or updates executed in (20). Table I details the global computational complexity of the detectors considered in this paper.

TABLE I
Computational complexity of detectors studied in this work.

Detector	Global Computational Complexity
AMP	$\mathcal{O}(N_\mathrm{I} N_\mathrm{t} N_\mathrm{r} + N_\mathrm{I} N_\mathrm{t}\sqrt{M})$
ZF	$\mathcal{O}(N_\mathrm{t}^3 + N_\mathrm{t}^2 N_\mathrm{r} + N_\mathrm{t} N_\mathrm{r})$
PDA	$\mathcal{O}(N_\mathrm{t}^4 + \sqrt{M}N_\mathrm{t}^3 + N_\mathrm{t}^2 N_\mathrm{r} + N_\mathrm{t} N_\mathrm{r})$
SD	$\mathcal{O}(M^{\sqrt{N_\mathrm{t}}})$
MLD	$\mathcal{O}(M^{N_\mathrm{t}})$

B. SER Performance

Here, we will consider two scenarios: i) square MIMO where $N_\mathrm{t} = N_\mathrm{r}$ and; ii) underloaded MIMO, where $N_\mathrm{t} < N_\mathrm{r}$.

1) Square MIMO Systems: Figures 1 (a) and (b) present the SER performance of all detectors considered in this paper for 16-QAM, assuming 4×4 and 8×8 MIMO, respectively.

From Figures 1 (a) and (b), we can see that the MLD and SD outperform significantly the remaining detectors. It is important to highlight that the PDA detector improves largely upon the performance of the ZF, which, as demonstrated in (13), is in fact used by the PDA as a preprocessing step in symbol detection. Furthermore, from Figures 1 (a) and (b) we conclude that the AMP suffers from poor performance in square MIMO systems. This behaviour potentially stems from the high inter antenna interference (IAI) present in square MIMO systems, which makes the distribution of the received signal depart from the typical independent identically distributed (iid) Gaussian. This is a crucial violation of the assumption made in operations carried out in (22).

2) Underloaded MIMO Systems: Figures 1 (c), (d) and (e) bring the SER performance considering the underloaded MIMO scenario, 16-QAM is considered here. The number of antennas is set to $N_\mathrm{r} = 64$ and $N_\mathrm{t} = 8, 16, 32$. In Figure 1 (c) we can see that the performance of all detectors significantly improves and the performance gap among them is reduced, with the PDA and ZF approaching the MLD performance and the AMP reaching the MLD performance for high SNR. The channel hardening is the main reason for this behaviour, since it makes the fading of MIMO channels behave deterministically [13]. This benefit is observed for the PDA detector also for 16 and 32 transmitting antennas, as can be seen in Figures 1 (d) and (e). As the number of transmitting antennas increases, ZF and AMP SER performances deteriorate considerably, to the point where the AMP detector presents an error floor. The AMP can only outperforms the ZF in the 32×64 MIMO for SNR between 9 dB and 14 dB, where the IAI is not prominent and the ZF noise enhancement is still high.

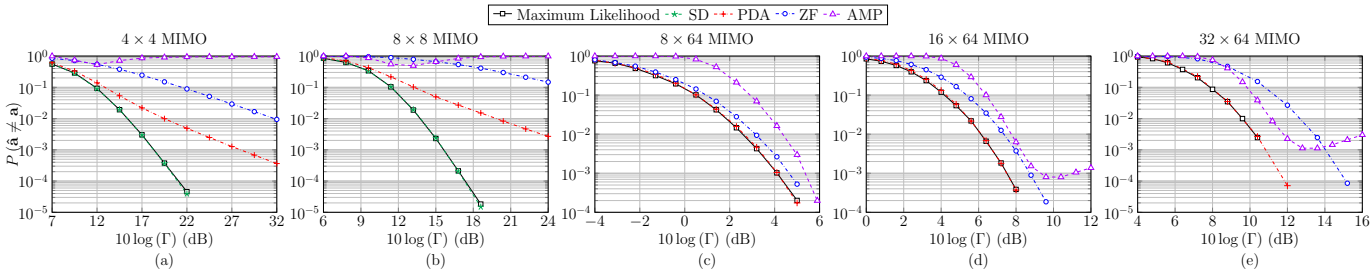

Fig. 1. Estimation of the detection performance for the MLD, SD, PDA, ZF and AMP detectors given in terms of SER ($P(\hat{\mathbf{a}} \neq \mathbf{a})$), for a range of SNR values. The square (a) 4×4 and (b) 8×8 and underloaded (c) 8×64, (d) 16×64 and (e) 32×64 MIMO systems are considered, also assuming a 16-QAM constellation.

TABLE II

THE COMPLEXITY-PERFORMANCE TRADE-OFF. THE TOPMOST ROW SHOWS THE HIGHEST ORDER TERM OF THE COMPLEXITY.

Complexity / Per. Δ	$(\cdot)^{N_t}$	N_t^4	N_t^3	N_t^2
≥ 3 (dB)	MLD/SD ($N_t = N_r$)	PDA ($N_t = N_r$)	—	—
$[1, 3)$ (dB)	MLD $(16, 32 \times 64)$	PDA $(16, 32 \times 64)$	ZF (8×64)	AMP (32×64)
< 1 (dB)	MLD (8×64)	PDA (8×64)	ZF (16×64)	—

C. Complexity-Performance Trade-off

Given the SER performance analysis done in Subsection IV-B and the computation complexity investigation presented in Subsection IV-A, it may be of interest that a joint complexity-performance analysis is produced comparing all detectors presented in this work. Table II brings the relative detection performance difference (Δ) obtained by a given detector and its next best, meaning that each cell of Table II is filled by a detector, if the other one closest in performance is so within any of the ranges expressed on the first column. It is important to mention that the trade-off analysis considers the region of values where $P(\hat{\mathbf{a}} \neq \mathbf{a}) \sim [10^{-3}, 10^{-2}]$. Note also that these detectors are ranked in descending order, from more to less complex as columns of Table II progress to the right.

From Table II, we can observe that detectors located near the upper-left corner yield greater gains in performance, but may be prohibitively complex. However, in specific scenarios, such as 8×64 and 32×64 MIMO, Table II shows that a good trade-off is achieved by the ZF and AMP detectors, respectively. Finally, note that for a heavily underloaded MIMO scenario, that is, 8×64, the PDA ceases to show an attractive trade-off, since the performance gap may be to small and the AMP, which is orders-of-magnitude less complex, can be used without significant losses in performance.

V. CONCLUSION

As devices limited in computational power start exploiting MIMO systems, the complexity of the detectors becomes an important aspect to be considered. Several MIMO detectors has been proposed recently aiming for presenting more attractive trade-off between complexity and SER performance. In this paper, we have analyzed such trade-off for 5 (five) detectors widely present in the literature, considering the complexity and SER performance for MIMO systems ranging from 4×4 up to 32×64 antennas. It shows that only the SD detector is able to achieve MLD performance for square MIMO. The PDA detector can achieve near-optimum performance in the underloaded scenarios, but the SER performance gap compared with the ZF might not justify the higher PDA complexity when $N_r >> N_t$. In this case, even the AMP can become an interesting choice, mainly for high SNR. Otherwise, the AMP is only interesting in very specific situations where the IAI is not the main limitation for the system SER performance and also when the ZF noise enhancement is high.

REFERENCES

[1] H. Tullberg, P. Popovski, Z. Li, M. A. Uusitalo, A. Hoglund, O. Bulakci, M. Fallgren, and J. F. Monserrat, "The METIS 5G System Concept: Meeting the 5G Requirements," *IEEE Commun. Mag.*, vol. 54, no. 12, pp. 132–139, 2016.

[2] Z. Chen, X. Zhang, S. Wang, Y. Xu, J. Xiong, and X. Wang, "Enabling Practical Large-Scale MIMO in WLANs With Hybrid Beamforming," *IEEE/ACM Trans. Net.*, vol. 29, no. 4, pp. 1605–1619, 2021.

[3] B. Hassibi and H. Vikalo, "On the sphere-decoding algorithm i. expected complexity," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2806–2818, 2005.

[4] S. Yang and L. Hanzo, "Fifty years of MIMO detection: The road to large-scale MIMOs," *IEEE Commun. Surveys Tutor.*, vol. 17, no. 4, pp. 1941–1988, 2015.

[5] M. A. M. Albreem and A. A. El-Saleh, "Approximate Matrix Inversion Methods vs. Approximate Message Passing (AMP) for massive MIMO Detectors," in *2019 IEEE 14th Malaysia International Conference on Communication (MICC)*, 2019, pp. 86–90.

[6] D. Pham, K. Pattipati, P. Willett, and J. Luo, "A generalized probabilistic data association detector for multiple antenna systems," *IEEE Commun. Lett.*, vol. 8, no. 4, pp. 205–207, 2004.

[7] B. Farhang-Boroujeny and H. Moradi, "OFDM Inspired Waveforms for 5G," *IEEE Commun. Surveys Tutor.*, vol. 18, no. 4, pp. 2474–2492, 2016.

[8] S. Yang, T. Lv, R. G. Maunder, and L. Hanzo, "From nominal to true a posteriori probabilities: An exact bayesian theorem based probabilistic data association approach for iterative MIMO detection and decoding," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 2782–2793, 2013.

[9] M. A. Albreem, M. Juntti, and S. Shahabuddin, "Massive MIMO detection techniques: A survey," *IEEE Commun. Surveys Tutor.*, vol. 21, no. 4, pp. 3109–3132, 2019.

[10] M. Varanasi, "Decision feedback multiuser detection: a systematic approach," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 219–240, 1999.

[11] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, 2009. [Online]. Available: https://www.pnas.org/doi/abs/10.1073/pnas.0909892106

[12] P. H. C. De Souza, L. L. Mendes, and M. Chafii, "Compressive learning in communication systems: A neural network receiver for detecting compressed signals in OFDM systems," *IEEE Access*, vol. 9, pp. 122 397–122 411, 2021.

[13] B. Hochwald, T. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1893–1909, 2004.