

A Low-Complexity Double EP-based Detector for Iterative Detection and Decoding in MIMO

Juan José Murillo-Fuentes, Irene Santos, José Carlos Aradillas and Matilde Sánchez-Fernández

Abstract—We propose a new iterative detection and decoding algorithm for multiple-input multiple-output (MIMO) based on expectation propagation (EP) with application to massive MIMO scenarios. Two main results are presented. We first introduce EP to iteratively improve the Gaussian approximations of both the estimation of the posterior by the MIMO detector and the soft output of the channel decoder. With this novel approach, denoted by *double-EP* (DEP), the convergence is very much improved with a computational complexity just two times the one of the linear minimum mean square error (LMMSE), as illustrated by the included experiments. Besides, as in the LMMSE MIMO detector, when the number of antennas increases, the computational cost of the matrix inversion operation required by the DEP becomes unaffordable. In this work we also develop approaches of DEP where the mean and the covariance matrix of the posterior are approximated by using the Gauss-Seidel and Neumann series methods, respectively. This low-complexity DEP detector has quadratic complexity in the number of antennas, i.e., the same as the low-complexity LMMSE techniques. Experimental results show that the new low-complexity DEP achieves the performance of the DEP as the ratio between the number of transmitting and receiving antennas decreases.

Index Terms—Expectation propagation, MMSE, low-complexity, iterative detection and decoding (IDD), massive MIMO, Gauss-Seidel, Neumann series.

I. INTRODUCTION

With multiple-input multiple-output (MIMO) strategies in wireless communication systems we may, among others, better face fading or provide an increased spectral efficiency, allowing for higher data rates [1], [2]. This technology is present in well-known standards such as LTE and Wi-Fi and has also been included in 5G. As the number of antennas is increased, larger multiplexing and/or diversity gains are attained. However, increasing the number of antennas poses challenging problems, one of them being the computational complexity associated to the detection of the transmitted symbols.

A MIMO detector provides an estimation of the transmitted symbols given the received signal and channel state information (CSI). This estimation is improved at the channel decoder. In iterative detection and decoding (IDD) schemes [3], [4], the output of the channel decoder is fed back to the MIMO

detector, further improving the estimation of the transmitted symbol through iterations.

For a large number of antennas, e.g. in massive MIMO [5]–[7], or high-order modulations, in LTE, 5G or Wi-Fi 6, optimal detection is unfeasible. The computation of the maximum a posteriori (MAP) or maximum likelihood (ML) criterion requires evaluating a huge number of probabilities to retain the largest one. Hence, its computational complexity grows exponentially with the number of antennas and the order of the constellation. In this scenario, approximate linear solutions, such as the zero forcing (ZF) or the linear minimum mean square error (LMMSE) [8], are a good tradeoff between complexity and performance [1], and also LMMSE based IDD has been proved to achieve capacity [9], [10].

In recent research it has been shown that expectation propagation (EP) [11]–[16] greatly outperforms LMMSE detection for high order discrete modulations, such as M -QAM for large M , increasing its computational complexity by just a few times, being of the same order. As developed in [11]–[15], an EP-based detector can be casted as a LMMSE algorithm where priors are replaced by some to be estimated Gaussian probability density function (pdf), given the observation. In particular, in [12], [14] EP is introduced to improve the detection performance for high-order constellations. EP is iteratively applied within the MIMO detector by using a damping approach and no feedback is provided from the decoder to the detector. These approaches assume perfect CSI. The extension of [14] to consider imperfect CSI is proposed in [13].

Despite EP and LMMSE approaches for $N_r \times N_t$ MIMO detection aim to reduce complexity from exponential to polynomial, both need to invert a $N_t \times N_t$ matrix, which requires a complexity of $\mathcal{O}(N_t^3)$, even if we resort to efficient inversion algorithms such as the Cholesky decomposition [4] or the Gauss-Jordan method [17]. This inversion is needed to obtain the mean and variance of the approximate posteriors at the output of the detector. Therefore, when the number of antennas increases, again the associated computational cost becomes excessively high.

In the literature we can find approximate detectors based on iterative methods to avoid the exact inverse computation, such as Neumann series expansion (NSE) [18]–[21], Gauss-Seidel (GS) [22], [23] or conjugate gradient (CG) [24]. Proposals in [18], [19] rely on NSE to compute an approximate inverse and then perform LMMSE detection and [21] approximates the inverse for an EP-based detector for massive MIMO. The GS-based LMMSE [23], hereafter denoted by LMMSE-GS, provides approximations to the mean values of the LMMSE by solving an associated linear system. It exhibits better

J.J. Murillo-Fuentes, I. Santos, and J.C. Aradillas are with the Dep. Teoría de la Señal y Comunicaciones, Universidad de Sevilla, Camino de los Descubrimientos s/n, 41092 Sevilla, Spain, E-mail: {murillo, irenesantos, jaradillas}@us.es. Matilde Sánchez-Fernández is with Dep. Teoría de la Señal y Comunicaciones, Universidad Carlos III de Madrid, Spain, E-mail: mati@tsc.uc3m.es.

Manuscript received X. This work was partially funded by Spanish government Spanish National Projects under grants TEC2017-90093-C3-2-R (TERESA-ADA) and TEC2016-78434-C3-R2 funded by (MINECO/AEI/FEDER, UE)

performance compared to NSE or CG in terms of convergence and complexity and for this reason we will use LMMSE-GS as a benchmark for comparison purposes. Up to our knowledge, GS has not been applied to EP MIMO detection, since it does not provide the estimation of the needed covariance matrix. Other approaches to approximate the inverse using EP are proposed in [25], [26], however, reported complexities scale as $\mathcal{O}(N_t^3)$.

All these low-complexity approaches focus on the approximate computation of the inverse, but furthermore they share the need for the Gram matrix to be computed in advance, involving in most of the cases a complexity that scales with $N_t^2 N_r$. Indeed, in [20] an NSE-based detection, hereafter denoted by LMMSE-NSE, is achieved with overall quadratic complexity in the number of antennas.

The EP-based techniques in [12]–[14], [23], [25], [26] describe solutions for standalone detection where EP is exclusively applied within the detector. A first version of EP-based IDD for MIMO is proposed in [11]. In the following this approach is denoted by EP-IC. In [15] the block expectation propagation (BEP) IDD is proposed by using damping as in [14] but redefining the EP parameters. The BEP exhibits much lower bit error rate (BER) than EP-IC. The computational complexity of the BEP is twice the one of the EP-IC and four times the one of the LMMSE. Both EP-based IDD approaches [11], [15] project into Gaussians pdf the information that is fed back from the channel decoder to the detector. EP was proposed in [27] to better approximate the information feedback from the channel decoder. In [16] this idea is exploited for MIMO. Similar approaches, not using EP but where some iterative approach is used in the detection and then feedback from the channel decoder is used to improve it, can be found in the literature [28].

In this work, it is first proposed to apply EP both in the detection and in the feedback from the channel decoder to the MIMO detector. The proposed detector iteratively applies EP algorithm twice: over an *inner loop* to obtain a more accurate posterior estimation at the output of the detector and within an *outer loop* to better approximate the discrete output of the channel decoder that is fed back to the detector to, in the next iteration, initialize the *inner loop*. This novel *double-EP* (DEP) MIMO IDD approach converges faster and reduces the number of iterations needed. As a result, it improves the BER of the BEP exhibiting the computational complexity of the EP-IC, i.e., twice the one of the LMMSE.

Second, low-complexity DEP approaches for large-scale MIMO, are explored. Scenarios with low ratios between the number of transmitting and receiving antennas, i.e., $\rho = N_t/N_r \ll 1$, are faced, where the matrices to be inverted are diagonally dominant and low-complexity approaches can be applied. It is highlighted that the accuracy in the estimation of the posterior mean has a greater impact in the performance. Therefore, the GS method is introduced to estimate it, as it achieves better approximation than the NSE or the CG. Besides, since this approach does not provide an estimation for the needed covariance matrix, NSE with two terms for its estimation is used. Furthermore, we found that for $\rho \ll 1$ the inner EP can be spared, while the EP in the outer loop

improves the BER at no extra computational cost. Accordingly, this new low-complexity DEP MIMO detector, denoted by approximate DEP based on GS (DEP-GS), has the same computational complexity order as the one of other LMMSE approximate methods.

By using the DEP-GS complexity scales as $\mathcal{O}(N_t^2)$ for the computation of the inverse, but still the Gram matrix is needed. To avoid the cubic dependence of the complexity of the Gram matrix with the number of antennas, a novel EP approach is also proposed borrowing from [20], the low-complexity DEP based on NSE (DEP-NSE), with overall complexity scaling with the square of the number of antennas.

In the included experiments, DEP outperforms BEP and EP-IC IDD in scenarios where matrices cannot be easily inverted, i.e., iterative low-complexity approximations for the inverse of the covariances matrices do not converge. As the channel matrix becomes orthogonal, the DEP-GS converges for $\rho \leq 1/4$, while the DEP-NSE achieves the BER of the DEP for $\rho \leq 1/8$. Remarkably, these low-complexity DEP proposals are able to achieve the performance of the DEP detector, even for high-order constellations.

II. SYSTEM MODEL

The information bit vector, \mathbf{a} , is encoded into the codeword vector, \mathbf{b} , which is the input to the modulator. Blocks of Q bits are interleaved and Gray-mapped into constellation \mathcal{A} , with $|\mathcal{A}| = 2^Q$ complex-valued symbols of mean energy E_s . The mapped symbols are demultiplexed into N_t substreams, one per antenna. The p th set of consecutive N_t symbols, $\mathbf{u}(p) = [u_1(p), \dots, u_{N_t}(p)]^T \in \mathbb{C}^{N_t \times 1}$, is sent over the channel. We denote by $b_{k,j}(p)$, with $j = 1, \dots, Q$, the bits mapped into the symbol transmitted by the k th antenna, $u_k(p)$. Hereafter, and for the sake of simplicity, index p will be omitted.

The received signal, $\mathbf{y} = [y_1, \dots, y_{N_r}]^T \in \mathbb{C}^{N_r \times 1}$, is given by

$$\mathbf{y} = \mathbf{H}\mathbf{u} + \mathbf{w} \quad (1)$$

where $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ is the channel matrix and $\mathbf{w} \sim \mathcal{CN}(\mathbf{w} : \mathbf{0}, \sigma_w^2 \mathbf{I}) \in \mathbb{C}^{N_r \times 1}$ is a circular complex-valued additive white Gaussian noise (AWGN) vector with, respectively, mean and covariance matrix $\boldsymbol{\mu}_w = \mathbf{0}$ and $\boldsymbol{\Sigma}_w = \sigma_w^2 \mathbf{I}$. The noise variance, σ_w^2 , and the channel matrix are known to the receiver. A soft detector computes the posterior probability of the transmitted symbol vector, \mathbf{u} , i.e.,

$$p(\mathbf{u} | \mathbf{y}, \mathbf{H}) \propto \mathcal{CN}(\mathbf{y} : \mathbf{H}\mathbf{u}, \sigma_w^2 \mathbf{I}) \prod_{k=1}^{N_t} p(u_k), \quad (2)$$

where the true prior, $p(u_k)$, with $k = 1, \dots, N_t$, can either be set to equiprobable distributions, $p(u_k) = 1/|\mathcal{A}|$, in a *standalone* detection, i.e., when no information from the channel decoder IDD is available, or can be set to the information fed back by the decoder in an IDD scenario.

Optimal standalone detectors need to compute all possible probabilities, $|\mathcal{A}|^{N_t}$. Note that these computations become unfeasible for large number of antennas and/or high-order modulations. Then, the soft detector outputs an extrinsic distribution that is demapped and fed to the decoder as extrinsic log-likelihood ratio (LLR), $L_E(b_{k,j})$, and the soft-decoding

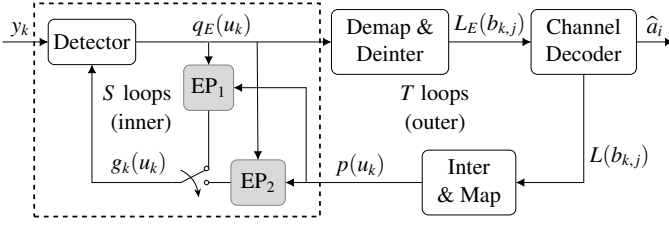


Fig. 1: Double EP-based IDD receiver diagram.

is performed. In IDD, the output of the channel decoder is furthermore mapped back to the detector to update the priors $p(u_k)$. This process is repeated iteratively for a given maximum number of iterations, T , or until convergence.

III. DOUBLE EP IDD

The proposed IDD technique, the DEP, implements a double loop where both detector and decoder apply EP, a Bayesian technique to estimate an approximation, within the exponential family, to a non-tractable distribution [29], [30].

The DEP performs two nested loops, see Fig. 1, one exclusively at the MIMO detector level (*inner loop*) and one comprising both the detector and decoder (*outer loop*). The inner loop (EP₁ in Fig. 1) is first used to approximate, after S iterations, the symbol posterior. Subsequently, the outer loop (EP₂ in Fig. 1) uses the extrinsic information at the output of the inner loop to soft-decode and further feed the soft-bit information back to the MIMO detector inner loop. The outer loop is run for T iterations.

A. DEP: Inner EP Loop

In MIMO, EP [14], [15] is applied to iteratively approximate the discrete priors in (2), $p(u_k)$, with Gaussians,

$$g_k^{[s]}(u_k) = \mathcal{CN}(u_k : \mu_{g_k}^{[s]}, \sigma_{g_k}^{2[s]}), \quad (3)$$

where $^{[s]}$ indicates the iteration number of the EP. This yields a Gaussian approximation to the posterior in (2),

$$q^{[s]}(\mathbf{u}) \propto \mathcal{CN}(\mathbf{y} : \mathbf{H}\mathbf{u}, \sigma_w^2 \mathbf{I}) \prod_{k=1}^{N_r} g_k^{[s]}(u_k) = \mathcal{CN}(\mathbf{u} : \boldsymbol{\mu}^{[s]}, \boldsymbol{\Sigma}^{[s]}) \quad (4)$$

where

$$\boldsymbol{\Sigma}^{[s]} = \left(\sigma_w^{-2} \mathbf{H}^H \mathbf{H} + \boldsymbol{\Lambda}_g^{[s]} \right)^{-1}, \quad (5)$$

$$\boldsymbol{\mu}^{[s]} = \boldsymbol{\Sigma}^{[s]} \left(\sigma_w^{-2} \mathbf{H}^H \mathbf{y} + \boldsymbol{\Lambda}_g^{[s]} \boldsymbol{\mu}_g^{[s]} \right), \quad (6)$$

$$\boldsymbol{\Lambda}_g^{[s]} = \text{diag} \left([1/\sigma_{g_1}^{2[s]}, \dots, 1/\sigma_{g_{N_r}}^{2[s]}] \right) \text{ and } \boldsymbol{\mu}_g^{[s]} = [\mu_{g_1}^{[s]}, \dots, \mu_{g_{N_r}}^{[s]}]^T.$$

Note that (4) resembles the LMMSE estimator but for the mean and covariance, that are iteratively updated by EP, given the observation.

Factors $g_k^{[s]}(u_k)$ are iteratively updated as follows in [15]. The extrinsic distributions are obtained,

$$q_E^{[s]}(u_k) = q^{[s]}(u_k) / g_k^{[s]}(u_k) = \mathcal{CN}(u_k : \mu_{E_k}^{[s]}, \sigma_{E_k}^{2[s]}), \quad (7)$$

Algorithm 1 Moment Matching (MM)

Given inputs: ϵ , $p(u_k)$ and $q_E(u_k)$,

- 1) Compute the moments $\mu_{\hat{p}_k}, \sigma_{\hat{p}_k}^2$ of $\hat{p}(u_k) = q_E(u_k)p(u_k)$. Set a *minimum allowed variance*, $\sigma_{\hat{p}_k}^2 = \max(\epsilon, \sigma_{\hat{p}_k}^2)$.
- 2) Run *moment matching*: set the mean and variance of the unnormalized Gaussian distribution

$$q_E(u_k) \cdot \mathcal{CN}(u_k : \mu_{g_k, \text{new}}, \sigma_{g_k, \text{new}}^2) \quad (9)$$

equal to $\mu_{\hat{p}_k}$ and $\sigma_{\hat{p}_k}^2$, to get the solution

$$\mu_{g_k, \text{new}} = (\sigma_{E_k}^2 \mu_{\hat{p}_k} - \sigma_{\hat{p}_k}^2 \mu_{E_k}) / (\sigma_{E_k}^2 - \sigma_{\hat{p}_k}^2), \quad (10)$$

$$\sigma_{g_k, \text{new}}^2 = \sigma_{\hat{p}_k}^2 \sigma_{E_k}^2 / (\sigma_{E_k}^2 - \sigma_{\hat{p}_k}^2). \quad (11)$$

Output: $\mu_{g_k, \text{new}}, \sigma_{g_k, \text{new}}^2$.

Algorithm 2 Moment Matching and Damping (EP₁)

Given inputs: β , $g_k^{[s]}(u_k)$, ϵ , $p(u_k)$ and $q_E^{[s]}(u_k)$,

- 1) Compute Alg. 1 (MM) with ϵ , $p(u_k)$ and $q_E^{[s]}(u_k)$ to provide $\mu_{g_k, \text{new}}^{[s]}, \sigma_{g_k, \text{new}}^{2[s]}$.
- 2) Run *damping*: update the values as

$$\sigma_{g_k}^{2[s+1]} = \left(\beta / \sigma_{g_k, \text{new}}^{2[s]} + (1 - \beta) / \sigma_{g_k}^{2[s]} \right)^{-1},$$

$$\mu_{g_k}^{[s+1]} = \sigma_{g_k}^{2[s+1]} \left(\beta \mu_{g_k, \text{new}}^{[s]} / \sigma_{g_k, \text{new}}^{2[s]} + (1 - \beta) \mu_{g_k}^{[s]} / \sigma_{g_k}^{2[s]} \right).$$

- 3) Control of *negative variances*:

if $\sigma_{g_k}^{2[s+1]} < 0$ **then** $\sigma_{g_k}^{2[s+1]} = \sigma_{g_k}^{2[s]}$, $\mu_{g_k}^{[s+1]} = \mu_{g_k}^{[s]}$.

Output: $\mu_{g_k}^{[s+1]}, \sigma_{g_k}^{2[s+1]}$.

where $q^{[s]}(u_k)$ are the marginals of the posterior in (4). Then, the EP computes $g_k^{[s+1]}(u_k)$ by matching the moments (MM) of the discrete posterior,

$$\hat{p}^{[s]}(u_k) = q_E^{[s]}(u_k)p(u_k), \quad (8)$$

with the ones of the new approximate posterior, $q_E^{[s]}(u_k)g_k^{[s+1]}(u_k)$. The MM is described in Alg. 1, where a minimum allowed variance, ϵ , is set. After MM, a damping procedure by means of parameter β is performed, checking for negative variances. Parameters β and ϵ control the convergence and instabilities of the algorithm. The whole procedure is summarized in Alg. 2. This algorithm is repeated over S iterations in the inner loop, see Fig. 1, to obtain the final extrinsic distributions, $q_E^{[s+1]}(u_k)$, that are sent to the channel decoder.

In the computation of (5) and (6) an $N_t \times N_t$ matrix is inverted. The initial computation of the Gram matrix $\mathbf{H}^H \mathbf{H}$ involves, approximately, $N_t^2 N_r / 2$ multiplications. The computational complexity of the inner EP loop (EP₁), including the cubic complexity of the matrix inversion, yields $O(N_t^2 N_r + (S+1)N_t^3)$.

B. DEP: Outer EP Loop

We denote by superindex $^{[t,s]}$ the s th iteration of the inner loop within the t th iteration of the outer loop. At the t th

Algorithm 3 DEP detector

Inputs: \mathbf{H} , σ_w^2 and \mathbf{y}

1) Initialization: set $p^{[0]}(u_k) = \frac{1}{|\mathcal{A}|} \sum_{u \in \mathcal{A}} \delta(u_k - u)$, $\epsilon = 10^{-8}$, $\mu_{g_k}^{[0,1]} = 0$ and $\sigma_{g_k}^{2[0,1]} = E_s$ for $k = 1, \dots, N_t$.

for $t = 0, \dots, T$ **do**

2) $\beta = \min(\exp(t/1.5)/10, 0.7)$.

EP at the inner loop:

for $s = 1, \dots, S$ **do**

3) Obtain $\Sigma_2^{[t,s]}$ and $\mu^{[t,s]}$ from (5)-(6) with $\mu_{g_k}^{[t,s]}$ and $\sigma_{g_k}^{2[t,s]}$.

for $k = 1, \dots, N_t$ **do**

4) Compute the k th extrinsic distribution, $q_E^{[t,s]}(u_k)$, as in (7).

5) Run Alg. 2 (EP₁) with β , $g_k^{[t,s]}(u_k)$, ϵ , $p^{[t]}(u_k)$ and $q_E^{[t,s]}(u_k)$ to obtain $\mu_{g_k}^{[t,s+1]}$, $\sigma_{g_k}^{2[t,s+1]}$.

end for

end for

6) Compute $q_E^{[t,S+1]}(u_k)$ as in (7), demap the extrinsic distribution and compute the extrinsic LLR, $L_E^{[t]}(b_{k,j})$.

7) Run the channel decoder to output $p^{[t+1]}(u_k)$.

EP at the outer loop:

if $t < T$ **then**

8) Run Alg. 1 (MM) with ϵ , $p^{[t+1]}(u_k)$ and $q_E^{[t,S+1]}(u_k)$ to obtain $\mu_{g_k}^{[t+1,1]}$, $\sigma_{g_k}^{2[t+1,1]}$.

9) Control of negative variances:

if $\sigma_{g_k}^{2[t+1,1]} < 0$ **then**

Set $\mu_{g_k}^{[t+1,1]} = \mathbb{E}_{p^{[t]}}[u_k]$ and $\sigma_{g_k}^{2[t+1,1]} = \mathbb{E}_{p^{[t]}}[(u_k - \mu_{g_k}^{[t+1,1]})^2]$.

end if

end if

end for

Output: Deliver $L_E^{[T]}(b_{k,j})$ to the decoder for $k=1, \dots, N_t$ and $j=1, \dots, Q$. Then provide the information bit estimates $\hat{\mathbf{a}}$.

iteration of the outer loop, after S iterations the inner loop provides the extrinsics $q_E^{[t,S+1]}(u_k)$ in (7) as input to the channel decoder. These extrinsics are used by the decoder to obtain $p^{[t+1]}(u_k)$, that are discrete probabilities. To map these probabilities back to the inner loop we need them to be Gaussian approximated. While in [11] these outputs are projected into Gaussians, here it is proposed to use EP instead, see EP₂ in Fig. 1. This EP₂ block uses Alg. 1 (MM) to match the moments of $\hat{p}^{[t+1]}(u_k) = q_E^{[t,S+1]}(u_k)p^{[t+1]}(u_k)$ and $q_E^{[t,S+1]}(u_k)g_k^{[t+1,1]}(u_k)$.

If the variances lead to negative values we replace them by the ones of the Gaussian projection. The resulting probabilities, $g_k^{[t+1,1]}(u_k)$, are used to initialize the inner EP.

The procedure is repeated over $t = 0, \dots, T$, with overall computational complexity of $\mathcal{O}(N_t^2 N_r + KN_t^3)$, where $K = (T+1)(S+1)$. The detailed implementation of this algorithm, referred to as DEP, is described in Alg. 3.

IV. GAUSS-SEIDEL AND NEUMANN METHODS FOR EP DETECTION

The mean and covariance matrix computation of the EP posterior in (5)-(6) lead to a computational complexity of

Algorithm 4 Low-complexity GS-NSE Detection

Inputs: \mathbf{A} , \mathbf{b} and P

1) Compute $\mathbf{D} = \text{diag}(\text{diag}(\mathbf{A}))$, $\mathbf{E} = \mathbf{A} - \mathbf{D}$ and $\mathbf{L} = \text{tril}(\mathbf{E})$.

2) Run NSE to obtain $\Sigma_2^{[t,s]}$ as in (13).

3) Compute the initial solution $\mu^{[t,s,(0)]}$ as in (15).

for $i = 1, \dots, P$ **do**

4) Run GS to obtain $\mu^{[t,s,(i)]}$ as in (14).

end for

Output: $\mu^{[t,s,(P)]}$ and $\Sigma_2^{[t,s]}$.

the DEP cubic in the number of antennas, due to the matrix inversion needed. To reduce it, we propose to develop low-complexity versions of the DEP based on the GS and NSE approaches. These methods converge if the matrix to invert, say \mathbf{A} , satisfies some characteristics. Specifically, the NSE requires \mathbf{A} to be diagonally dominant, while GS needs \mathbf{A} to be Hermitian positive definite. In our case of study, at a given iteration of the DEP, the inverse of $\mathbf{A} = \sigma_w^{-2} \mathbf{H}^H \mathbf{H} + \Lambda_g^{[t,s]}$ yields (5). In large-scale MIMO with $N_r \gg N_t$, the columns of \mathbf{H} are asymptotically orthogonal [1], [31], which guarantees that the Gram matrix $\mathbf{H}^H \mathbf{H}$ is symmetric positive definite and diagonally dominant. Since σ_w^2 is positive and $\Lambda_g^{[t,s]}$ is a diagonal matrix with positive entries, \mathbf{A} is also symmetric positive definite and diagonally dominant. This fact enables the convergence of both NSE and GS methods when applied to MIMO detection.

These techniques use \mathbf{A} expressed as $\mathbf{A} = \mathbf{D} + \mathbf{E}$, where $\mathbf{D} = \text{diag}(\text{diag}(\mathbf{A}))$ is a diagonal matrix with the diagonal entries of \mathbf{A} . The NSE can approximate the inversion of \mathbf{A} as [19]

$$(\mathbf{A})^{-1} \approx \sum_{n=0}^{R-1} \left(-(\mathbf{D})^{-1} \mathbf{E} \right)^n (\mathbf{D})^{-1}, \quad (12)$$

where R indicates the number of terms of (12). Note that if $R \rightarrow \infty$, then (12) yields the exact inverse. The computational complexity of (12) is of $\mathcal{O}(N_t^2)$ for $R \leq 2$ and $\mathcal{O}(N_t^3)$ (or higher) for $R > 2$. For this reason, hereafter it is used (12) with $R = 2$ to approximate the matrix inversion in (5), yielding

$$\Sigma_2^{[t,s]} \approx \Sigma_2^{[t,s]} = (\mathbf{D})^{-1} - (\mathbf{D})^{-1} \mathbf{E} (\mathbf{D})^{-1}. \quad (13)$$

Also, the GS method solves a linear system of equations, $\mathbf{A} \mathbf{x} = \mathbf{b}$, that can be used to obtain an approximation of the mean, bypassing the need of matrix inversion. We exploit this technique to estimate $\mu^{[t,s]}$ in (6) by setting $\mathbf{x} = \mu^{[t,s]}$, $\mathbf{A} = \sigma_w^{-2} \mathbf{H}^H \mathbf{H} + \Lambda_g^{[t,s]}$ and $\mathbf{b} = \sigma_w^{-2} \mathbf{H}^H \mathbf{y} + \Lambda_g^{[t,s]} \mu_g^{[t,s]}$. The GS iteratively estimates it as

$$\mu^{[t,s,(i)]} = (\mathbf{D} + \mathbf{L})^{-1} \left(\mathbf{b} - (\mathbf{L})^H \mu^{[t,s,(i-1)]} \right) \quad (14)$$

where $i = 1, \dots, P$, with P the number of iterations of the GS algorithm and $\mathbf{L} = \text{tril}(\mathbf{E})$, i.e., the lower triangular part of \mathbf{E} . The computational complexity of (14) is dominated by the inversion of a lower triangular matrix that requires $\mathcal{O}(N_t^2)$ [22].

Here it is proposed to use GS to better estimate the mean values of the approximation of the EP posterior, keeping in

Algorithm 5 Low-complexity DEP-GS IDD

Inputs: P , \mathbf{H} , σ_w^2 and \mathbf{y}

Run Alg. 3 with Step 3 replaced by:

3) Run Alg. 4 with $\mathbf{A} = \sigma_w^{-2} \mathbf{H}^H \mathbf{H} + \mathbf{\Lambda}_g^{[t,s]}$, $\mathbf{b} = \sigma_w^{-2} \mathbf{H}^H \mathbf{y} + \mathbf{\Lambda}_g^{[t,s]} \boldsymbol{\mu}_g^{[t,s]}$ and P , to obtain $\boldsymbol{\mu}^{[t,s]} = \boldsymbol{\mu}^{[t,s],(P)}$ and $\boldsymbol{\Sigma}^{[t,s]} = \boldsymbol{\Sigma}_2^{[t,s]}$.

Output: Deliver $L_E^{[T]}(b_{k,j})$ to the decoder for $k=1, \dots, N_r$ and $j=1, \dots, Q$. Then provide the information bit estimates $\hat{\mathbf{a}}$.

mind that errors on the approximation of the mean have a great impact on the BER. Since GS does not provides the low-complexity estimation of the inverse of the matrix, just solves the associated linear system, we propose to use NSE with two terms for the covariance matrix estimation. Hence, the overall complexity scales as $\mathcal{O}(N_r^2)$.

GS initialization in (14), $\boldsymbol{\mu}^{[t,s],(0)}$, has an important role in its convergence. Also, it has a direct impact on the computational complexity, since the faster the convergence is achieved, the lower the value of P needed. Following the guidelines in [23], we exploit the NSE-based inverse approximation in (12) with $R = 2$ to obtain an initial value for the GS iteration process,

$$\boldsymbol{\mu}^{[t,s],(0)} = \boldsymbol{\Sigma}_2^{[t,s]} \mathbf{b}. \quad (15)$$

The whole procedure of the proposed GS-NSE approximation of the mean and covariance matrix of the EP posterior in (5)-(6) is described in Alg. 4. When this approximation is combined with the DEP explained in Section III, it yields the low-complexity DEP-GS detector. Its implementation is detailed in Alg. 5.

The overall complexity, that includes the computation of (15), scales as $\mathcal{O}(N_r^2 N_r + K(P+2)N_r^2)$, where typically $P \leq 3$. Note that the second term of the computational cost is significantly lower than KN_r^3 in the exact DEP.

At this point it is important to remark that DEP-GS with $P = 0$, $S = 0$ and $T = 0$ is the approximate LMMSE detection via NSE with two terms proposed in [18], [19]. Also, if $P > 0$, $S = 0$ and $T = 0$ then it yields the LMMSE-GS in [23]. The DEP-GS with $P = 0$, $S > 0$ and $T = 0$ is the approximate EP standalone detection via NSE with two terms proposed in [21]. Finally, the DEP-GS with $P = 0$ and $T = 0$ yields an improved version, with the results in [15], of the NSE-based EP method proposed in [21]. The EP method in [21] needs $S = 20$, involving a high computational complexity. Besides, by numerical simulation we found that, as expected, the DEP-GS with $P = 1$ or $P = 2$ improves its performance, with the same complexity for the inverse computation, scaling as $\mathcal{O}(N_r^2)$. Accordingly, we do not include [21] in the comparisons. In [20] it is proposed an LMMSE NSE-based MIMO detection different to the one described in (12). The authors introduce further approximations for the covariance matrix and rewrite (12) to avoid the initial computation of the Gram matrix. We extend the proposal in [20] to DEP by replacing the a priori mean and covariance matrix of the symbol transmitted in [20] with, respectively, the EP mean and covariance matrix approximations $\boldsymbol{\mu}_g^{[t,s]}$ and $(\mathbf{\Lambda}_g^{[t,s]})^{-1}$. We use NSE therein to

TABLE I: Order of the computational complexities.

Algorithm	Complexity	Complexity in Sec. Section V
LMMSE	$N_r^2 N_r + T' N_r^3$	$N_r^2 N_r + T' N_r^3$
EP-IC [11]	$N_r^2 N_r + T' 2 N_r^3$	$N_r^2 N_r + T' 2 N_r^3$
BEP [15]	$N_r^2 N_r + K N_r^3$	$N_r^2 N_r + T' 4 N_r^3$
DEP	$N_r^2 N_r + K N_r^3$	$N_r^2 N_r + T' 2 N_r^3$
LMMSE-GS [23]	$N_r^2 N_r + T'(P+2)N_r^2$	$N_r^2 N_r + T' 4 N_r^2$
DEP-GS	$N_r^2 N_r + K(P+2)N_r^2$	$N_r^2 N_r + T' 4 N_r^2$
DEP-NSE-I	$K(4+2R)N_r N_r$	$T' 10 N_r N_r$
DEP-NSE-II	$N_r^2 N_r + K(P+2)N_r^2$	$N_r^2 N_r + T' 4 N_r^2$

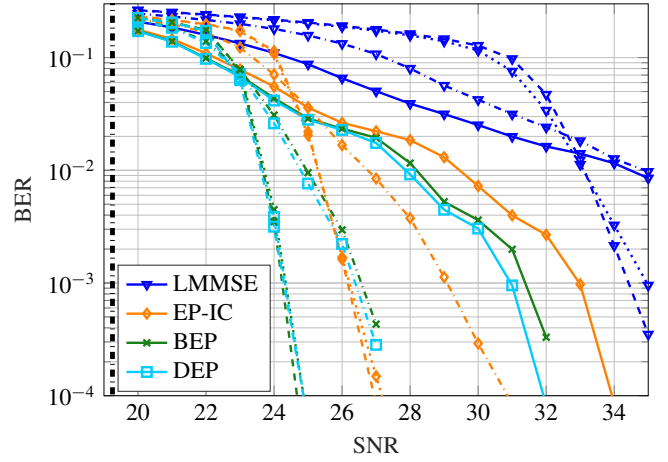


Fig. 2: A comparison of the BER vs. SNR for growing $N_r = N_r$: 4 (solid), 8 (dash-dotted), 64 (dashed) and 128 (dotted) for the LMMSE, the EP-IC with $S=1$ [11], BEP with $S=3$ [15] and DEP with $S=1$ with a 256-QAM.

estimate the covariance matrix in (5) and the mean in (6). We will refer to this approach as DEP-NSE-I. The computational complexity of DEP-NSE-I scales as $\mathcal{O}(K(4+2R)N_r N_r)$ where R is the number of iterations of the NSE. This DEP-NSE can be run, with the same result, changing the order in the multiplication of matrices, leading to the complexity order of the GS approximations. This other development is denoted hereafter by DEP-NSE-II. Following the guidelines in [15], the proposed parameters the DEP, DEP-GS and DEP-NSE are set to $\epsilon = 10^{-8}$ and $\beta = \min(\exp(t/1.5)/10, 0.7)$.

The computational complexity of the proposed algorithms and the benchmark algorithms is detailed in Tab. I. In the second column the complexity is given in terms of the tunable parameters for a trade-off between performance and complexity: $K = (T+1)(S+1)$, $T' = T+1$, the number of iterations of the GS algorithm, P , and the number of terms of the NSE approximation, R . The third column (Complexity in Sec. Section V) reflects the complexity given the parametrization used in next section, in terms of T' and replacing the values of inner iterations, S , P and R used in the experiments.

V. EXPERIMENTAL RESULTS

In this section we analyze the performance of IDD strategies with the proposed DEP (\square), its low-complexity versions, DEP-GS (\circ) and DEP-NSE ($*$), LMMSE (∇), LMMSE-GS with $P = 2$ [23] (\triangle) and the inner loop EP approaches BEP [15]

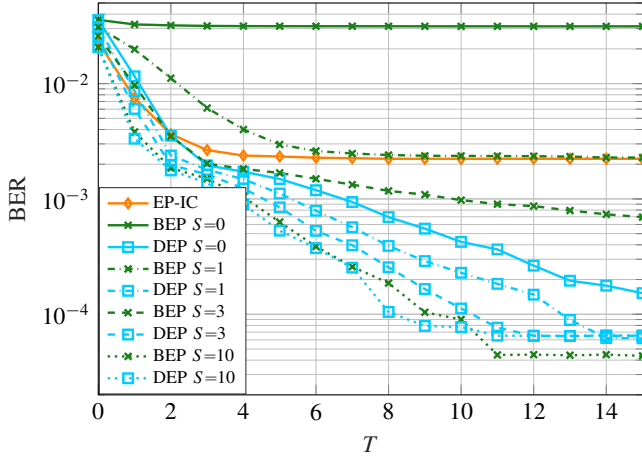


Fig. 3: BER versus the number of outer iterations T at SNR=30 dB, with a 256-QAM, 6×6 antennas and $S = 0, 1, 3, 10$, for the BEP and DEP. EP-IC BER is also depicted.

(\times) and EP-IC [11] (\diamond). The number of iterations for the DEP-GS and DEP-NSE were set to $P = 2$ and $R = 3$ [20]. Unless otherwise indicated, $S = 1$ for the EP-IC, $S = 3$ for the BEP, $S = 1$ for the DEP and $S = 0$ for the DEP-GS and DEP-NSE. The value of $S = 1$ for DEP leads to the same computational complexity as EP-IC and doubles the one of the LMMSE. These parameters were selected to minimize the BER while reducing the computational complexity, after intensive simulation studies and in view of the experiments shown in this section. In Tab. I, third column, we include the order of the computational complexity for the methods in these experiments.

We depict the BER for several scenarios. In the experiments, unless otherwise indicated, results are averaged over 10^2 random encoded words of length $V = 4096$ using low-density parity-check (LDPC) codes of rate $r = 1/2$, sent through 10^2 random channel samples. The number of outer iterations was set to $T = 10$. Each complex valued channel coefficient is independent and identically Gaussian distributed with zero mean and unit variance. The absolute value of LLRs given to the decoder is limited to 3 in order to avoid very confident probabilities. In some of the figures, the BER at the waterfall for the SIMO case with equal ρ as in the given experiment is depicted as reference, in dashed thick black line.

A. Non-orthogonal MIMO Scenarios

In an uncorrelated MIMO scenario, as N_t tends to N_r , the Gram matrix $\mathbf{H}^H \mathbf{H}$ loses the property of orthogonality between columns. Other sources of MIMO orthogonality loss in $\rho = N_t/N_r \ll 1$ is spatial correlation in the propagation scenario, due either both to low scattering or compact antenna designs, and here again the Gram matrix $\mathbf{H}^H \mathbf{H}$ is not diagonal dominant anymore. To analyze the effect of this orthogonality loss in DEP and its low-complexity approaches we focus in MIMO scenarios where $\rho = N_t/N_r \ll 1$ with no spatial correlation. In Fig. 2 we depict the BER versus SNR for *square* ($\rho = 1$) configurations, i.e., $N = N_t = N_r$ systems, and some of the proposed solutions. We did not depict the

approximations, as they exhibited poor results in the square case. As N grows, from 4 to 128, the DEP and BEP perform equally good. For large N , $N = 64$ and $N = 128$, the solutions provide the same BER. Here the EP-IC exhibits a good performance but 2 dB far from the DEP/BEP solutions. As the number of antennas decreases, the probability of the channel matrix to be singular increases as well. Accordingly, for a low number of antennas the BER increases and DEP/BEP approaches exhibit even larger gains with respect to the EP-IC.

The computational complexity of the DEP, with $S = 1$, is half the one of BEP with $S = 3$ as in [15], and similar to the one of EP-IC, with $S = 1$, see Tab. I. In Fig. 3 we include the convergence versus the number of outer iterations T with different values of S for the BEP and the DEP in the $N_t = N_r = 6$ scenario. Note that the BEP with $S = 0$ yields the LMMSE. It can be observed how with just the outer EP, for $S = 0$, the DEP outperforms the BEP with $S = 3$. Also, with $S = 1$, the DEP exhibits a good performance in terms of BER while the computational complexity is half the one of the BEP.

Fig. 4 includes the study of the state evolution [10], [16] for the DEP focusing on different points of the IDD scheme. We compute the MSE between the true values of the symbols and the ones estimated at the output of the EP based detector (EP₁ in our proposed IDD scheme) (\circ), the output of the channel decoder (input to the EP₂ in our proposed IDD scheme) (\diamond) and the output of a LMMSE detector whose priors have been initialized with the output values of the channel decoder, or equivalently, in our proposed IDD scheme, the input to the EP detector (EP₁) (\square). The constellation, a 256-QAM, is normalized to unit energy. It can be observed, for different SNR and from top to bottom, how the estimations are being refined. We set $S = 10$ to better observe the improvement of the inner EP loop.

The proposed solutions present a robust behaviour regardless of the rate of the code. In Fig. 6 we depict the BER for

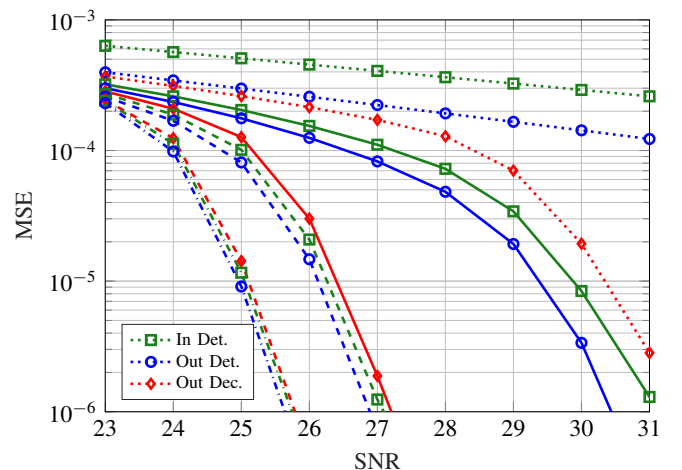
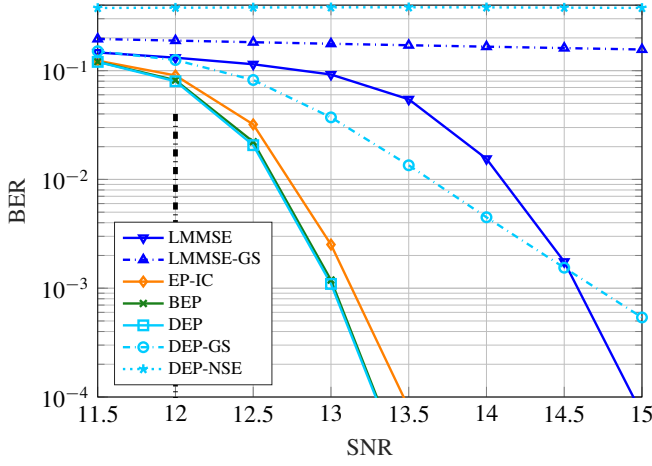
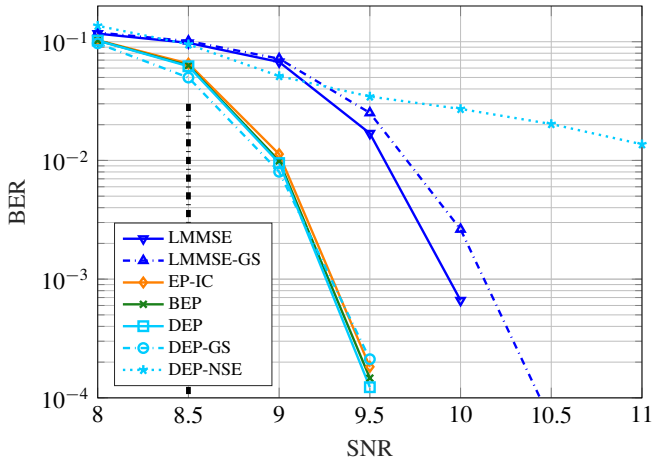


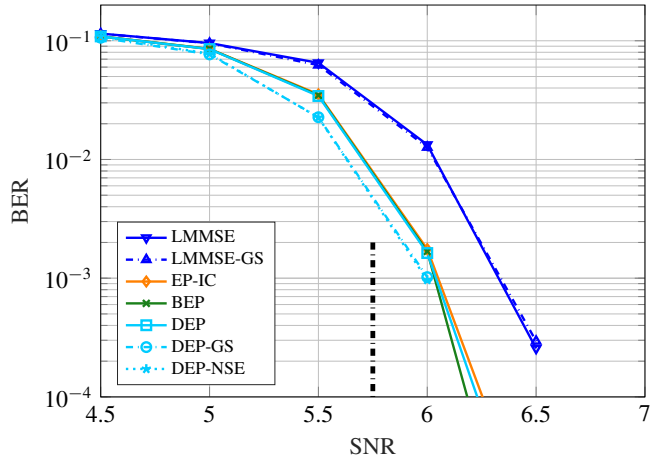
Fig. 4: MSE vs. SNR for the DEP in the 128×128 antennas and 256-QAM scenario after $T = 1$ (dotted), $T = 2$ (solid), $T = 3$ (dashed) and $T = 4$ (dash-dotted) iterations at the output of the channel decoder, the input to the detector and the output of the detector.



(a)



(b)



(c)

Fig. 5: BER for a) 16×32 , b) 16×64 and c) 16×128 antennas, 64-QAM, with LMMSE and its GS approximation with $P=2$, EP-IC with $S=1$ [11], BEP with $S=3$ [15], DEP with $S=1$ and $T=10$ and DEP-GS, DEP-NSE approximations with $S=0$ and $T=10$.

$N = 128$ as in Fig. 2 but with larger code words, $V = 64800$, and rates $r = 1/4, 1/3, 1/2$ and $3/4$. As the rate increases, the impact of the EP-based detector in the inner loop is emphasized, with higher gains with respect to the EP-IC.

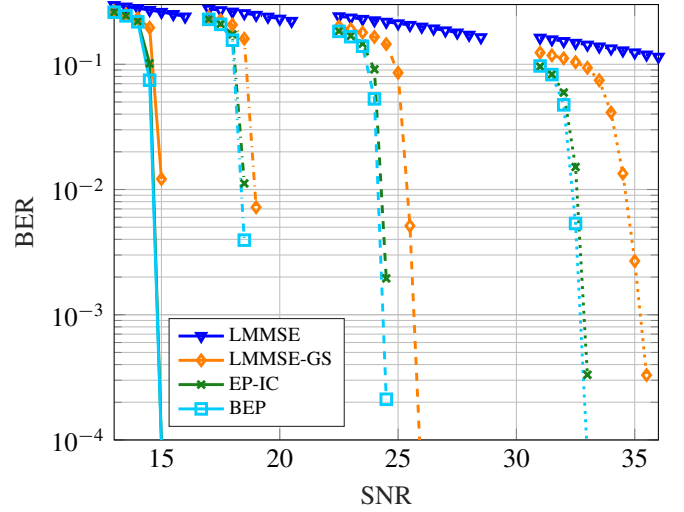


Fig. 6: BER vs. SNR in the 128×128 antennas scenario for $V = 64800$ and rates $1/4$ in (solid), $1/3$ (dash-dotted), $1/2$ (dashed) and $3/4$ (dotted) for the LMMSE, the EP-IC, DEP and BEP. A 256-QAM was transmitted and $T = 5$.

For $N_t < N_r$ MIMO channel matrices with $\rho = \frac{1}{2} \ll 1$, see Fig. 5.a and Fig. 8.a, low-complexity techniques might enhance with respect to the squared case but are still far from the solution of the DEP, BEP or EP-IC. It can be observed that for $\rho = 1/2$, DEP-GS provides some useful output just for the 64-QAM scenario. Besides, BEP and DEP exhibit a similar performance, slightly better than the one of the EP-IC.

B. Orthogonal MIMO Scenarios

As ρ decreases the columns of the Gram matrix $\mathbf{H}^H \mathbf{H}$ become more diagonal, since the columns of \mathbf{H} are more orthogonal, favouring detection. In Fig. 5.b-c and Fig. 8.b-c where $\rho \leq 1/4$ the DEP, BEP or EP-IC achieve the same BER, with 0.5 dB gain compared to the LMMSE. Furthermore, low-complexity approaches can be used as conditions for their convergence are met. It can be observed in Fig. 5.b and Fig. 8.b

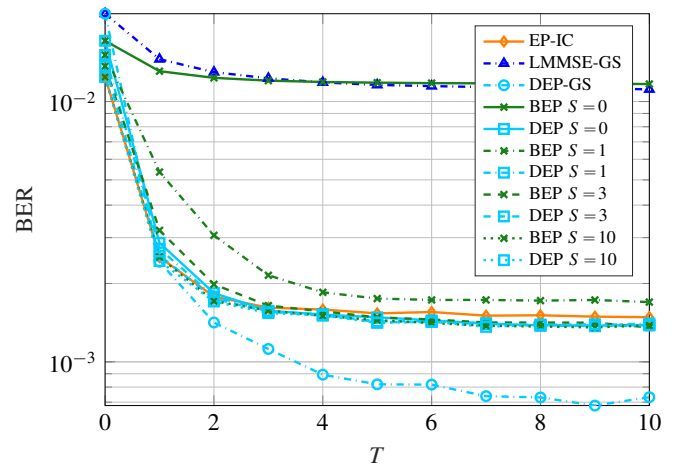
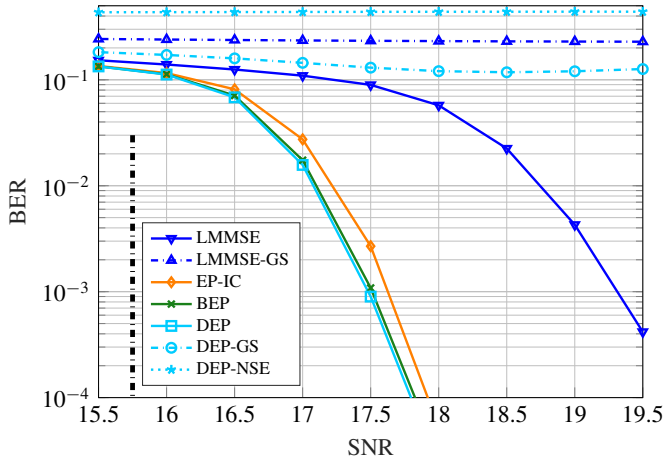
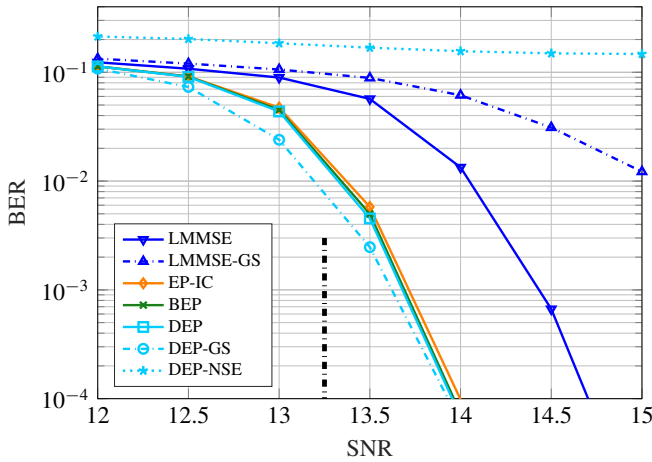


Fig. 7: BER versus the number of outer iterations T at $\text{SNR}=6$ dB, for a 64-QAM and 16×128 antennas and $S = 0, 1, 3, 10$, for the BEP and DEP. EP-IC BER is also depicted.

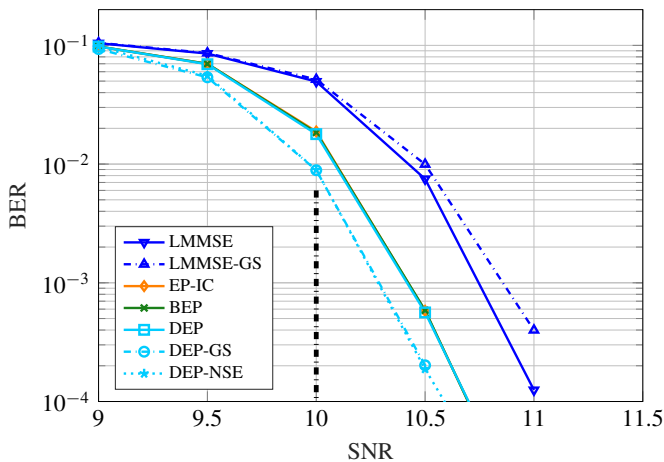
how the IDD with the DEP-GS detector achieves a good result, even slightly better than the one of the DEP. Note that in previous works, see [20, Fig. 1] where $T = 0$, this effect can be observed, here is further enhanced by the IDD process. As



(a)



(b)



(c)

Fig. 8: BER for a) 16×32 , b) 16×64 and c) 16×128 antennas, 256-QAM, with LMMSE and its GS approximation with $P=2$, EP-IC with $S=1$ [11], BEP with $S=3$ [15], DEP with $S=1$ and $T=10$ and DEP-GS, DEP-NSE approximations with $S=0$ and $T=10$.

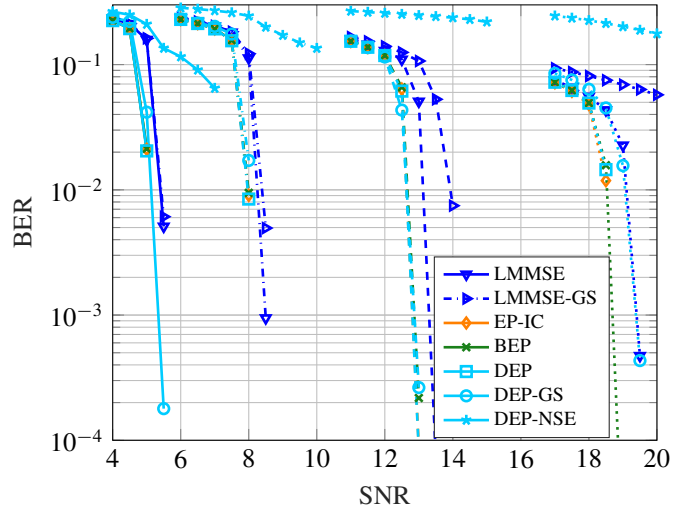


Fig. 9: BER vs. SNR in the 16×64 antennas scenario for $V = 64800$ and rates $1/4$ in (solid), $1/3$ (dash-dotted), $1/2$ (dashed) and $3/4$ (dotted) for the LMMSE, LMMSE-GS, the EP-IC, DEP, BEP, and DEP-GS and DEP-NSE. A 256-QAM was transmitted and $T = 5$.

ρ decreases, for $\rho = 1/8$ in Fig. 5.c and Fig. 8.c the DEP-NSE provides the same BER as the one of the DEP-GS with a much lower complexity.

The proposed solutions present a robust behaviour regardless of the rate of the code. In Fig. 9 we depict the BER for the same scenario as in Fig. 8.b, i.e., 16×64 antennas, $V = 64800$ bits as code length and rates $r = 1/4, 1/3, 1/2$ and $3/4$. As the rate decreases, the feedback from the channel decoder very much reduces the BER. For low values, even the LMMSE-GS provides quite a good result. As the rate is increased, e.g., $r = 3/4$, the LMMSE-GS becomes unavailable, while non approximating solutions exhibit a gain of 0.5 dB.

In Fig. 7 we depict the BER versus the number of outer iterations. It can be concluded that as the Gram matrix becomes diagonal, the DEP provides approximately the same BER regardless of the number of inner iterations S . Accordingly, we used $S = 0$ for the low-complexity EP approaches. This BER is similar to the BER of the EP-IC. The BEP needs $S > 3$ to achieve this error. Besides, the DEP-GS method has not the lowest BER for $T = 0$, however, after a couple of outer iterations it has slightly outperformed the other approaches.

C. Computational complexity

First, it is important to remark that EP itself involves a computational complexity linear in the number of unknowns. Hence, the complexity of these novel proposals is not driven by the EP technique, but by the computation of the inverse of the matrix associated to the MIMO detection in the inner loop. While in a LMMSE stand-alone solution we need to invert a matrix once, in the inner loop of our IDD scheme, DEP and BEP need $S + 1$ inversions to refine the solution with EP. Hence, the DEP and the BEP involve $S + 1$ times the equivalent computational complexity of computing a linear MMSE. In the outer loop, in the feedback from the channel decoder, there is no inversion associated and the computational complexity of the EP is again linear with the number of variables.

This is evidenced in Tab. I, where we present the complexity for the parameter values used in this section. LMMSE has complexity order of N_t^3 with the number of outer iterations, T' , while BEP, with $S = 3$, has $4N_t^3$. The EP-IC has $2N_t^3$ but worse performance, see the square case in the experiments. In this paper we investigate a new detector with the performance of the BEP and the computational complexity of the EP-IC, i.e., $2N_t^3$. Introducing EP in the outer loop quite improves the BEP, allowing for similar performance with much less inner iterations. Note that the computational complexity order of the DEP in Table I, for $S = 1$, is twice the one of the LMMSE.

If conditions are good enough, for $\rho < 1$, inner iterations could be avoided, with $S = 0$. For $S = 0$ the DEP has the same computational complexity as the LMMSE. In Fig. 7 it can be observed that DEP solutions for $S = 0$ to $S = 10$ in (□) are quite similar. The BEP would need $S \geq 1$ to achieve the performance of the DEP. In addition, the approximations to compute the inverse can be introduced, further reducing the computational complexity, as in Fig. 5 or Fig. 8, where the DEP with $S = 0$ and approximations for the inverse are analyzed. In these scenarios, the DEP-NSE-II in Tab. I might have a lower computational complexity than the DEP-NSE-I.

VI. CONCLUSIONS

The novel DEP IDD MIMO detector applies the EP algorithm twice: to obtain a more accurate Gaussian approximation at the output of the detector and to better approximate with Gaussians the discrete output at the channel decoder. This novel detector greatly outperforms the LMMSE and the EP-IC approaches with a complexity that is just two times the one of the LMMSE. However, this complexity, that scales with N_t^3 , can be excessively high for large number of antennas. To avoid this drawback, we also propose a low-complexity DEP detector based on the NSE and GS methods to approximate, respectively, the covariance matrix and the mean of the EP solution. In the experimental results included, the low-complexity DEP approaches the BER of the DEP as the ratio ρ decreases, with quadratic complexity in the number of antennas.

Although the DEP with $S = 1$ is presented as a robust solution with a good compromise between complexity and performance, in scenarios where detection is not challenging, S can be set to zero, with similar complexity as the LMMSE. Besides, approximate solutions can also be applied, with further reduced computational complexity and similar performance.

REFERENCES

- [1] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan 2013.
- [2] R. W. Heath and A. Lozano, *Foundations of MIMO Communication*. Cambridge University Press, 2018.
- [3] R. E. Chall, F. Nouvel, M. Hélar, and M. Liu, "Iterative receivers combining MIMO detection with turbo decoding: performance-complexity trade-offs," *EURASIP J. Wireless. Commun. and Netw.*, no. 1, pp. 1–19, Mar 2015.
- [4] C. Studer, S. Fateh, and D. Seethaler, "ASIC implementation of soft-input soft-output MIMO detection using MMSE parallel interference cancellation," *IEEE J. Solid-State Circuits*, vol. 46, no. 7, pp. 1754–1765, July 2011.
- [5] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, February 2014.
- [6] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics in Signal. Process.*, vol. 8, no. 5, pp. 742–758, Oct 2014.
- [7] S. Yang and L. Hanzo, "Fifty years of MIMO detection: The road to large-scale MIMOs," *IEEE Commun. Surveys Tutorials*, vol. 17, no. 4, pp. 1941–1988, 2015.
- [8] L. Liu, C. Yuen, Y. L. Guan, Y. Li, and Y. Su, "Convergence analysis and assurance for gaussian message passing iterative detector in massive MU-MIMO systems," *IEEE Trans. Wireless Comm.*, vol. 15, no. 9, pp. 6487–6501, 2016.
- [9] L. Liu, Y. Chi, C. Yuen, Y. L. Guan, and Y. Li, "Capacity-achieving MIMO-NOMA: Iterative LMMSE detection," *IEEE Trans. Signal Process.*, vol. 67, no. 7, pp. 1758–1773, 2019.
- [10] L. Liu, C. Liang, J. Ma, and L. Ping, "Capacity optimality of AMP in coded systems," 2019.
- [11] M. Senst and G. Ascheid, "How the framework of expectation propagation yields an iterative IC-LMMSE MIMO receiver," in *Proc. IEEE Global Telecomm. Conf. (GLOBECOM)*, Dec 2011, pp. 1–6.
- [12] J. Céspedes, P. M. Olmos, M. Sánchez-Fernández, and F. Pérez-Cruz, "Expectation propagation detection for high-order high-dimensional MIMO systems," *IEEE Trans. Commun.*, vol. 62, no. 8, pp. 2840–2849, Aug 2014.
- [13] K. Ghavami and M. Naraghi-Pour, "MIMO detection with imperfect channel state information using expectation propagation," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 8129–8138, Sept 2017.
- [14] J. Céspedes, P. M. Olmos, M. Sánchez-Fernández, and F. Pérez-Cruz, "Probabilistic MIMO symbol detection with expectation consistency approximate inference," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3481–3494, April 2018.
- [15] I. Santos and J. J. Murillo-Fuentes, "Self and turbo iterations for MIMO receivers and large-scale systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 1095–1098, Aug 2019.
- [16] J. Ma, L. Liu, X. Yuan, and L. Ping, "On orthogonal AMP in coded linear vector systems," *IEEE Trans. Wireless Comm.*, vol. 18, no. 12, pp. 5658–5672, 2019.
- [17] J. Arias-García, R. P. Jacobi, C. H. Llanos, and M. Ayala-Rincón, "A suitable FPGA implementation of floating-point matrix inversion based on Gauss-Jordan elimination," in *VII Southern Conf. on Programmable Logic (SPL)*, April 2011, pp. 263–268.
- [18] M. Wu, B. Yin, A. Vosoughi, C. Studer, J. R. Cavallaro, and C. Dick, "Approximate matrix inversion for high-throughput data detection in the large-scale MIMO uplink," in *IEEE Int. Symp. Circuits and Syst. (ISCAS)*, May 2013, pp. 2155–2158.
- [19] M. Wu, B. Yin, G. Wang, C. Dick, J. R. Cavallaro, and C. Studer, "Large-scale MIMO detection for 3GPP LTE: Algorithms and FPGA implementations," *IEEE J. Sel. Topics in Signal Processing*, vol. 8, no. 5, pp. 916–929, Oct 2014.
- [20] L. Fang, L. Xu, and D. D. Huang, "Low complexity iterative MMSE-PIC detection for medium-size massive MIMO," *IEEE Wireless Commun. Lett.*, vol. 5, no. 1, pp. 108–111, Feb 2016.
- [21] Y. Zhang, Z. Wu, C. Li, Z. Zhang, X. You, and C. Zhang, "Expectation propagation detection with Neumann-series approximation for massive MIMO," in *2018 IEEE Int. Work. Signal Process. Syst. (SIPS)*, Oct 2018, pp. 59–64.
- [22] L. Dai, X. Gao, X. Su, S. Han, C. I, and Z. Wang, "Low-complexity soft-output signal detection based on Gauss-Seidel method for uplink multiuser large-scale MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp. 4839–4845, Oct 2015.
- [23] C. Zhang, Z. Wu, C. Studer, Z. Zhang, and X. You, "Efficient soft-output Gauss-Seidel data detector for massive MIMO systems," *IEEE Trans. Circuits and Syst. I: Regular Papers*, pp. 1–12, 2018.
- [24] B. Yin, M. Wu, J. R. Cavallaro, and C. Studer, "Conjugate gradient-based soft-output detection and precoding in massive MIMO systems," in *IEEE Global Commun. Conf.*, Dec 2014, pp. 3696–3701.
- [25] X. Tan, Y. Ueng, Z. Zhang, X. You, and C. Zhang, "A low-complexity massive MIMO detection based on approximate expectation propagation," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 7260–7272, 2019.
- [26] G. Yao, G. Yang, J. Hu, and C. Fei, "High-efficiency expectation propagation detector for high-order massive MIMO systems," *IEEE Access*, vol. 7, pp. 125 225–125 239, 2019.
- [27] J. Hu, H. Loeliger, J. Dauwels, and F. Kschischang, "A general computation rule for lossy summaries/messages with examples from equalization," in *Proc. 44th Allerton Conf. Comm., Control, and Comput.*, Sep 2006, pp. 27–29.

- [28] Y. Yang, W. Wang, and X. Gao, "AMP dual-turbo iterative detection and decoding for LDPC coded multibeam MSC uplink," *China Communications*, vol. 15, no. 6, pp. 178–186, 2018.
- [29] T. Minka, "Expectation propagation for approximate Bayesian inference," in *Proc. 17th Conf. Uncertainty Art. Int. (UAI)*, 2001, pp. 362–369.
- [30] M. Seeger, "Expectation propagation for exponential families," *Univ. Calif., Berkeley, CA, USA, Tech. Rep.*, 2005.
- [31] X. Gao, L. Dai, Y. Ma, and Z. Wang, "Low-complexity near-optimal signal detection for uplink large-scale MIMO systems," *Electro. Lett.*, vol. 50, no. 18, pp. 1326–1328, August 2014.