



HAL
open science

Low-Complexity Scheduled Expectation Propagation based on QRD and SIC

Adam Mekhiche, Antonio Maria Cipriano, Charly Poulliat

► **To cite this version:**

Adam Mekhiche, Antonio Maria Cipriano, Charly Poulliat. Low-Complexity Scheduled Expectation Propagation based on QRD and SIC. IEEE Communications Letters, 2023, 27 (10), pp.1-1. 10.1109/LCOMM.2023.3303610 . hal-04181069

HAL Id: hal-04181069

<https://hal.science/hal-04181069>

Submitted on 12 Sep 2023

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

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

Low-Complexity Scheduled Expectation Propagation based on QRD and SIC

Adam Mekhiche, Antonio Maria Cipriano, Charly Poulliat

Abstract—This letter presents novel Expectation Propagation (EP) algorithms for Single User (SU) and Multiple Users (MU) Multiple-Input Multiple-Output (MIMO) detection that address the performance loss and slow convergence of classic Scalar EP (SEP) solutions with even lower computational complexity. The proposed algorithms employ QR Decomposition (QRD), optimized EP message scheduling, and Successive Interference Cancellation (SIC) techniques to achieve superior performance and convergence speed comparable to those of Vector Expectation Propagation (VEP). Our results are illustrated through error rate curves and analysis of performance-complexity trade-offs in SU and MU scenarios.

Index Terms—Expectation Propagation, QR Decomposition, Scheduling, Successive Interference Cancellation

I. INTRODUCTION

MIMO with Orthogonal Frequency Division Multiplexing (OFDM) is a widely used technology at the core of many wireless telecommunication standards like Wireless Fidelity (WiFi), 4G - Long Term Evolution (LTE) and 5G - New Radio (NR), in part, due to the possible high spectral efficiency gain through Spatial Multiplexing (SM). One recent proposal is to boost performance by introducing tens of antennas, e.g. massive MIMO (mMIMO), which comes at the cost of high computational complexity for detection, i.e. an exponential increase with the number of antennas. Bit Interleaved Coded Modulation with Iterative Detection (BICM-ID) can achieve optimal performance through the iterative exchange of soft information, usually based on Log-Likelihood Ratios (LLR), between a Soft-Input Soft-Output (SISO) detector and a SISO decoder at the cost of an even higher complexity due to several instances of detection and decoding. To cope with the increase in complexity, new SISO detection algorithms have been proposed based on Message Passing (MP), such as Belief Propagation (BP) [1], Approximate Message Passing (AMP) [2], or Vector AMP (VAMP) [2], [3] algorithms. These MP Algorithms (MPA) aim at computing a low-complexity approximation of the symbol-wise Maximum A Posteriori (MAP) criterion, which is too costly in practice.

In this paper, we focus on MPA derived within the EP framework [4], which is another approach to deriving an efficient MPA. EP has been shown to be a good candidate for MIMO and mMIMO applications compared to other MPA. BP

has a computational complexity slightly higher than MAP [1], while VAMP requires a costly Singular Value Decomposition (SVD) for a matrix inversion. Authors from [5] proposed to apply EP on a scalar Factor Graph (FG) to reduce the complexity, by removing the need to invert a full matrix, compared to EP used on a vector FG like in [6], [7]. Similarly to [8], authors from [9] proposed a QRD pre-processing with EP to reduce the number of messages exchanged during an EP iteration and improve performance. Even though better performance is achieved, there is still room for improvement to close the gap between the performance of a SEP [5], [9], [10] and a VEP [6], [7], [10].

In this paper, we propose an optimized scheduling of the exchange of EP messages to improve both convergence speed and performance for SU-MIMO and MU-MIMO. The article is organized as follows: In Section II, the system model is introduced, and then, the proposed scheduled SEP with QRD is explained in Section III. The related algorithm is detailed in Section III-A for the SU setting and in Section III-B for the MU setting. Its complexity is assessed in Section III-C and compared with other competitive MPA. Finally, simulation results are presented in Section IV.

II. SYSTEM MODEL

A spatially multiplexed $N_t \times N_r$ SU-MIMO-OFDM system using a Space-Time (ST) BICM-ID scheme [11] with N_t transmitting antennas and N_r receiving antennas is considered. At time $n \in \llbracket 1, N_s \rrbracket$, a stream of information bits $\mathbf{b}_n \in \mathbb{F}_2^{K_b}$, where \mathbb{F}_2 is the finite field of order 2, is encoded using an Error Correcting Code (ECC) of length N and rate $R = K_b/N$. The encoded bits $\mathbf{c}_n \in \mathbb{F}_2^N$ are interleaved with a random interleaver, and then Gray-mapped to symbols of a constellation $\mathcal{X} \subset \mathbb{C}$. A symbol x represents a tuple of m bits and there are $|\mathcal{X}| = M = 2^m$ symbols in the constellation. The mapping function is denoted $\varphi : \mathbb{F}_2^m \rightarrow \mathcal{X}$ while $\varphi^{-1} : \mathcal{X} \rightarrow \mathbb{F}_2^m$ is the demapping function that returns the m -tuple of coded bits. $\forall k \in \llbracket 1, m \rrbracket, \varphi_k^{-1} : \mathcal{X} \rightarrow \mathbb{F}_2$ is the demapping function that returns the single k^{th} bit corresponding to a symbol x . The vector of symbols $\mathbf{x}_n = [x_{n,1}, \dots, x_{n,N_t}]$ is then splitted across the N_t antennas. This system model employs a Space-Time (ST) [11] Bit-Interleaved Coded Modulation with Iterative Decoding (BICM-ID) scheme. The other model studied is single antenna MU-MIMO where, instead of a single information bits stream, there are $N_u = N_t$ user and one stream per user, which is equivalent to N_u SU-MIMO. At the time n , the signal \mathbf{x}_n is sent through N_t transmitting antennas on a non-correlated ergodic Rayleigh channel $\mathbf{H}_n \in \mathbb{C}^{N_r \times N_t}$

This work has been partly funded by the French National Research Agency project EVASION, grant reference ANR-20-CE25-0008-01.

A. Mekhiche is with Thales, 92230 Gennevilliers, France and with IRIT-INPT, University of Toulouse, 31000 Toulouse, France.

A.M. Cipriano is with Thales, 92230 Gennevilliers, France.

C. Poulliat is with IRIT-INPT, Toulouse University, 31000 Toulouse, France. email: name.surname@thalesgroup.com or name.surname@toulouse-inp.fr

and the signal $\mathbf{y}_n \in \mathbb{C}^{N_r}$ is received using N_r antennas. The corresponding linear model is:

$$\mathbf{y}_n = \mathbf{H}_n \mathbf{x}_n + \mathbf{w}_n, \quad (1)$$

with $\mathbf{w}_n \in \mathbb{C}^{N_r}$ an Additive White Gaussian Noise (AWGN) sample with the properties $\mathbb{E}(\mathbf{w}_n) = 0$ and $\mathbb{E}(\mathbf{w}_n \mathbf{w}_n^H) = N_0 \mathbf{I}_{N_r}$. The time index will be omitted for readability except when needed for comprehension. The Channel State Information (CSI) available on the receiver side is considered perfect. The turbo-iterated receiver aims at computing the probability $\mathbb{P}(\mathbf{x}|\mathbf{y}, \mathbf{H}, N_0)$ which can be factorized as:

$$\mathbb{P}(\mathbf{x}|\mathbf{y}, \mathbf{H}, N_0) \propto \prod_{j=1}^{N_r} \underbrace{\mathbb{P}(y_j|\mathbf{x}, \mathbf{H}, N_0)}_{f_j^{\text{EQU}}} \prod_{i=1}^{N_t} \underbrace{\mathbb{P}(x_i|\mathbf{c}_i)}_{f_i^{\text{DEM}}} \underbrace{\mathbb{P}(\mathbf{c}_i)}_{f_i^{\text{DEC}}} \quad (2)$$

where f_j^{EQU} are likelihood functions, f_i^{DEM} are the factors associated with the demodulation process, and f_i^{DEC} are priors associated with coded binary vector labels \mathbf{c}_i , which are provided by the ECC during the turbo-iterated process.

From the factorization (2), the corresponding Factor Graph (FG) can be represented as in Fig. 1, which is the typical scalar FG in [5], [9], [10]. This FG is the support on which Message Passing Algorithms (MPAs) are applied to approximate the Probability Density Function (PDF) $\mathbb{P}(\mathbf{x}|\mathbf{y}, \mathbf{H}, N_0)$. Authors from [5], [9] applied Expectation Propagation (EP) [4] on this FG. EP is an approximate Bayesian inference technique within the message-passing framework. It tracks a distribution p by approximating it with another distribution q from a specific set \mathcal{Q} , from the exponential family. To project p on the set \mathcal{Q} , i.e. choosing the statistically closest distribution $q \in \mathcal{Q}$, EP uses the inclusive Kullback-Leibler divergence [12]:

$$D_{KL}(p||q) = \int p(x) \log \left(\frac{p(x)}{q(x)} \right) dx \quad (3)$$

so the projection can be written :

$$q = \text{proj}_{\mathcal{Q}}[p] = \arg \min_{\tilde{q} \in \mathcal{Q}} D_{KL}(p||\tilde{q}). \quad (4)$$

EP exchanges messages, which are Gaussian distributions, along the edges of an FG, and the message from a factor node f_j to a variable node x_i is:

$$m_{f_j \rightarrow x_i}(x_i) = \frac{1}{m_{x_i \rightarrow f_j}(x_i)} \times \text{proj}_{\mathcal{Q}} \left[m_{x_i \rightarrow f_j}(x_i) \int_{\mathbf{x}} f_j(\mathbf{x}) \prod_{\substack{i' \in \mathcal{N}(f_j) \\ i' \neq i}} m_{x_{i'} \rightarrow f_j}(x_{i'}) \right]. \quad (5)$$

with $\mathcal{N}(f_j)$ the set of connected variable nodes to (f_j) . To improve EP on this FG, authors from [9] proposed a QR Decomposition as a pre-processing of the received signal, which helps reduce the number of edges in the graph and improves performance. This matrix decomposition was also studied with other MPAs like Belief Propagation [8], [13]. Applying QRD on \mathbf{H} , i.e. $\mathbf{H} = \mathbf{Q}\mathbf{R}$, gives the equivalent observation model

$$\tilde{\mathbf{y}} = \mathbf{Q}^H \mathbf{y} = \mathbf{R} \mathbf{x} + \tilde{\mathbf{w}} \quad (6)$$

with $\mathbf{Q} \in \mathbb{C}^{N_r \times N_r}$, a unitary matrix, $\mathbf{R} \in \mathbb{C}^{N_r \times N_t}$, an upper triangular matrix and $\tilde{\mathbf{w}} = \mathbf{Q}^H \mathbf{w}$. Statistical properties of the additive noise remain unchanged due to the unitary property of

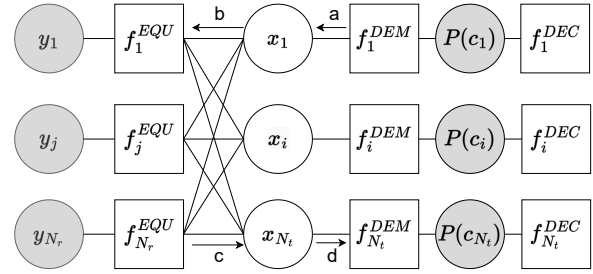


Fig. 1. Factor Graph representation of (2) with $a \triangleq \mathcal{CN}(\overleftarrow{\mu}_i^d, \overleftarrow{\nu}_i^d)$, $b \triangleq \mathcal{CN}(\overleftarrow{\mu}_{i,j}^e, \overleftarrow{\nu}_{i,j}^e)$, $c \triangleq \mathcal{CN}(\overleftarrow{\mu}_{i,j}^e, \overleftarrow{\nu}_{i,j}^e)$, $d \triangleq \mathcal{CN}(\overleftarrow{\mu}_i^d, \overleftarrow{\nu}_i^d)$.

Q. The new FG after QR pre-processing is shown in Fig. 2. In the next section, we propose improved versions of EP applied to a scalar FG, which shows better performance and faster convergence than existing SEP algorithms.

III. PROPOSED SCHEDULED SCALAR EXPECTATION PROPAGATION WITH QRD

A. SSEP QRD in SU-MIMO

Authors from [5], [7], [9], [10] proposed a flooding scheduling of the EP messages on the graph. The proposed algorithm aims at computing $\mathbb{P}(\mathbf{x}|\mathbf{y}, \mathbf{H})$ more efficiently than the classic Scalar EP (SEP) [5] or QRD SEP [9], thanks to better scheduling of the message exchange. The QRD pre-processing of the channel matrix generates a new sparser FG (Fig. 2), which can be more efficiently browsed in the same manner as a classic Block Decision Feedback Equalizer (BDFE) [14].

The algorithm starts the messages exchange at the bottom of the FG to compute an estimation of x_{N_t} , which is connected to all the equalizing nodes. It means, $\forall j \in \llbracket 1, N_r \rrbracket$, it propagates the messages $m_{f_j^{\text{EQU}} \rightarrow x_{N_t}}(x_{N_t})$ to gather all the available information about x_{N_t} at the equalizing side of the FG. Instead of simultaneously propagating messages toward the others x_i like in SEP [5] or QRD SEP [9], the following propagated message is $m_{x_{N_t} \rightarrow f_{N_t}^{\text{DEM}}}(x_{N_t})$. Then, information coming from the demapper is sent to the variable node x_{N_t} with the message $m_{f_{N_t}^{\text{DEM}} \rightarrow x_{N_t}}(x_{N_t})$.

At this point, the variable x_{N_t} has been estimated with all the available information. Hence, the following targeted variable node is x_{N_t-1} . It will be more accurately estimated than in a flooding scheduling since the equalizing nodes have a better estimation of x_{N_t} , as it comes from the demapper. Messages are exchanged similarly, propagating the estimation to the demapper and sending the demapping information to the equalizing nodes, up to node x_1 . The previous process can be eventually repeated multiple times (denoted as *auto-iterations*). After L auto-iterations, the messages $m_{x_i \rightarrow f_j^{\text{EQU}}}(x_i)$ are propagated once more time. Compared to [9], this step results in a slightly higher complexity, which is analyzed in Section III-C. The overall number of equalization steps and demapping steps is the same as in QRD SEP. This algorithm is presented in Alg. 1 and can be used for SU- or MU-MIMO. More efficient scheduling is also possible for MU-MIMO and SU-MIMO with V-BLAST [15]. It will be presented in the next section.

Algorithm 1 Scheduled Scalar EP with QRD

Input: $\mathbf{y}, \mathbf{H}, N_0, T = \text{Turbo iteration}, L = \text{Auto iteration}$
Output: $\hat{\mathbf{b}}$ - the estimated bits.

```

1:  $\mathbf{H} = \mathbf{QR}$ ,  $\mathbf{y} \leftarrow \mathbf{Q}^H \mathbf{y}$ ,  $\mathbf{H} \leftarrow \mathbf{R}$ 
2:  $\forall i, k \in \llbracket 1, N_t \rrbracket, \llbracket 1, m \rrbracket, \lambda_{i,k}^a = 0$ 
3: for  $t = 0 : T$  do
4:    $\forall i, j \in \llbracket 1, N_t \rrbracket, \llbracket 1, N_r \rrbracket,$ 
5:    $m_{f_j^{\text{EQU}} \rightarrow x_i}(x_i) = m_{x_i \rightarrow f_i^{\text{DEM}}}(x_i) = \mathcal{CN}(0, +\infty)$ 
6:   for  $i = 1 : N_t$  do
7:     Compute  $m_{f_i^{\text{DEM}} \rightarrow x_i}(x_i)$ :
8:      $\tilde{q}(x_i) \propto \exp\left(-\frac{|x_i - \mu_i^d|^2}{\nu_i^d} - \sum_{k=1}^m \varphi_k^{-1}(x_i) \lambda_{i,k}^a\right)$ 
9:      $q(x_i) \sim \mathcal{CN}(\mu_i^d = \mathbb{E}[\tilde{q}(x_i)], \nu_i^d = \text{Var}[\tilde{q}(x_i)])$ 
10:     $\overleftarrow{\mu}_i^d = \overleftarrow{\nu}_i^d \left( \frac{\mu_i^d}{\nu_i^d} - \frac{\overrightarrow{\mu}_i^d}{\overrightarrow{\nu}_i^d} \right)$  and  $\overleftarrow{\nu}_i^d = \left( \frac{1}{\nu_i^d} - \frac{1}{\overrightarrow{\nu}_i^d} \right)^{-1}$ 
11:   end for
12:   for  $i = 1 : N_t$  and then for  $j = 1 : N_r$  do
13:     Compute  $m_{x_i \rightarrow f_j^{\text{EQU}}}(x_i)$ :
14:      $\overleftarrow{\nu}_{i,j}^e = \left( (\nu_i^d)^{-1} - (\overrightarrow{\nu}_{i,j}^e)^{-1} \right)^{-1}$ 
15:      $\overleftarrow{\mu}_{i,j}^e = \overleftarrow{\nu}_{i,j}^e \left( \mu_i^d / \nu_i^d - \overrightarrow{\mu}_{i,j}^e / \overrightarrow{\nu}_{i,j}^e \right)$ 
16:   end for
17:   for  $l = 0 : L$  and then  $i = N_t : -1 : 1$  do
18:     for  $j = 1 : N_r$  do
19:       Compute  $m_{f_j^{\text{EQU}} \rightarrow x_i}(x_i)$ :
20:        $\overrightarrow{\nu}_{i,j}^e = \left( N_0 + \sum_{i' \neq i} |h_{j,i'}|^2 \overleftarrow{\nu}_{i',j}^e \right) / |h_{j,i}|^2$ 
21:        $\overrightarrow{\mu}_{i,j}^e = \left( y_j - \sum_{i' \neq i} h_{j,i'} \overleftarrow{\mu}_{i',j}^e \right) / h_{j,i}$ 
22:     end for
23:     Compute  $m_{x_i \rightarrow f_i^{\text{DEM}}}(x_i)$ :
24:      $\overrightarrow{\nu}_i^d = \left( \sum_{j' \in \mathcal{N}(x_i)} \frac{1}{\overrightarrow{\nu}_{i,j'}^e} \right)^{-1}$ ,  $\overrightarrow{\mu}_i^d = \overrightarrow{\nu}_i^d \left( \sum_{j' \in \mathcal{N}(x_i)} \frac{\overrightarrow{\mu}_{i,j'}^e}{\overrightarrow{\nu}_{i,j'}^e} \right)$ 
25:     if  $l \neq L$  then
26:       Compute  $m_{f_i^{\text{DEM}} \rightarrow x_i}(x_i)$ :
27:       Mean and variance as in Alg. 1 line 7.
28:       for  $j = 1 : N_r$  do
29:         Compute  $m_{x_i \rightarrow f_j^{\text{EQU}}}(x_i)$ :
30:         Mean and variance as in Alg. 1 line 13.
31:       end for
32:     end if
33:   end for
34:   for  $i = 1 : N_t$  and then for  $k = 1 : m$  do
35:      $\lambda_{i,k}^e = \log\left( \frac{\sum_{x_i \in \mathcal{X}: \varphi_k^{-1}(x_i)=1} \tilde{q}(x_i)}{\sum_{x_i \in \mathcal{X}: \varphi_k^{-1}(x_i)=0} \tilde{q}(x_i)} \right) - \lambda_{i,k}^a$ 
36:   end for / end for
37:   Send  $\lambda^e$  to decoder and receive  $\lambda^a$  and  $\hat{\mathbf{b}}$ 
38: end for

```

B. SSEP QRD SIC for MU-MIMO

In MU-MIMO, it is possible to benefit from the error correction capability of the ECC within a turbo-iteration through the use of Successive Interference Cancellation (SIC). Indeed, we propose another detector using a Scheduled Scalar Expectation Propagation with QR Decomposition and SIC (QRD SSEP SIC). Instead of propagating messages of each user up to the demapper, we propose to go through the decoder and use the extrinsic output of the decoder to refine the soft estimation of the associated symbol.

Fig. 2 shows the differences between QRD SSEP in SU-MIMO, where decoding is done concurrently since a single codeword is split across all antennas, and QRD SSEP SIC in MU-MIMO, where each user encodes its codeword which

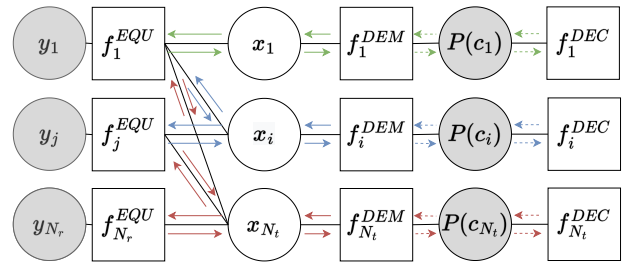


Fig. 2. Factor Graph representation of (2) after QRD. Red messages are propagated first, then the blue ones, and finally, the green ones. Dotted lines are done in parallel for SU-MIMO and serialized, in accordance with their color, for MU-MIMO.

Algorithm 2 Scheduled Scalar EP with QRD and SIC

Input: $\mathbf{y}, \mathbf{H}, N_0, T = \text{Turbo iteration}, L = \text{Auto iteration}$

Output: $\hat{\mathbf{b}}$ - the estimated bits.

```

1:  $\mathbf{H} = \mathbf{QR}$ ,  $\mathbf{y} \leftarrow \mathbf{Q}^H \mathbf{y}$ ,  $\mathbf{H} \leftarrow \mathbf{R}$ 
2:  $\forall i, k \in \llbracket 1, N_t \rrbracket, \llbracket 1, m \rrbracket, \lambda_{i,k}^a = 0$ 
3: for  $t = 0 : T$  do
4:    $\forall i, j \in \llbracket 1, N_t \rrbracket, \llbracket 1, N_r \rrbracket,$ 
5:    $m_{f_j^{\text{EQU}} \rightarrow x_i}(x_i) = m_{x_i \rightarrow f_i^{\text{DEM}}}(x_i) = \mathcal{CN}(0, +\infty)$ 
6:   for  $i = 1 : N_t$  do
7:     Compute  $m_{f_i^{\text{DEM}} \rightarrow x_i}(x_i)$  as in Alg. 1 line 7.
8:   end for
9:   for  $i = 1 : N_t$  and then for  $j = 1 : N_r$  do
10:    Compute  $m_{x_i \rightarrow f_j^{\text{EQU}}}(x_i)$  as in Alg. 1 line 13.
11:   end for
12:   for  $l = 0 : L$  and then  $i = N_t : -1 : 1$  do
13:     for  $j = 1 : N_r$  do
14:       Compute  $m_{f_j^{\text{EQU}} \rightarrow x_i}(x_i)$  as in Alg. 1 line 19.
15:     end for
16:     Compute  $m_{x_i \rightarrow f_i^{\text{DEM}}}(x_i)$  as in Alg. 1 line 23.
17:     if  $l = L$  then
18:       Compute  $\lambda_i^e$  as in Alg. 1 line 35.
19:       Send  $\lambda^e$  to decoder and receive  $\lambda^a$  and  $\hat{\mathbf{b}}$ 
20:     end if
21:     Compute  $m_{f_i^{\text{DEM}} \rightarrow x_i}(x_i)$  as in Alg. 1 line 7.
22:     for  $j = 1 : N_r$  do
23:       Compute  $m_{x_i \rightarrow f_j^{\text{EQU}}}(x_i)$  as in Alg. 1 line 13.
24:     end for
25:   end for
26: end for

```

can be independently decoded. This scheduling enables the refinement of the detection process for the other users by obtaining a more accurate estimation of a user's symbol after a single decoding step. It eliminates the need to wait for all users to be decoded, i.e. the next turbo-iteration, to converge.

C. Complexity Analysis of SSEP QRD

In this section, a precise estimation of the computational complexity is performed to assess the gain of the proposed algorithm.

VEP [6] has an $\mathcal{O}(N^3)$ complexity as it requires a $N \times N$ matrix inversion, with $N = \min(N_t, N_r)$, to compute messages coming from the equalizer node. DEP [7] has the same complexity as VEP since both use the same FG and apply the same EP message rule on it. Their difference lies in saving messages $m_{x_i \rightarrow f_i^{\text{DEM}}}(x_i), \forall i \in \llbracket 1, N_t \rrbracket$, from one turbo iteration to the next for DEP, and resetting these messages for VEP. Authors of VEP also proposed to keep the same EP messages in [16, Chapter 4, Section 3]. SEP [5] does

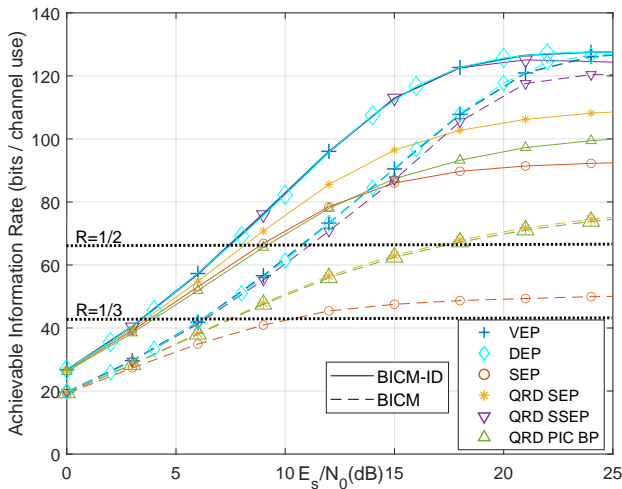


Fig. 3. Achievable Information Rate (AIR) of a 32×32 SU-MIMO with a 16QAM ST-BICM and ST-BICM-ID transmissions.

not require any matrix inversion, and its overall complexity is smaller than VEP even though there are more messages to be exchanged between equalizing nodes and variable nodes x_i , $N_t \times N_r$ instead of N_t . Its most computationally demanding messages are the messages from the equalizing nodes to variable node x_i . QRD SEP [9] must perform one QR decomposition per MIMO symbol, no matter the number of auto and turbo iterations, and QRD has a complexity $\mathcal{O}(N_t^3/3)$. Once this pre-processing is done, the messages are a bit less complex, fewer sums in (19), and there are $N_t(N_t + 1)/2$ messages instead of $N_t \times N_r$. Finally, the proposed QRD SSEP has a slightly higher complexity than QRD SEP since there are $N_t(N_t + 1)/2$ more messages from variable node x_i to equalizing node f_j^{EQU} . This additional complexity is minor compared to the overall complexity: it results in a 5% or less increase in complexity for QRD SSEP compared to QRD SEP.

Table I shows two relative computational complexities per MIMO scenario. The first percentage is the relative complexity of QRD SSEP over VEP to highlight the gain of the proposed algorithms compared to the optimal EP. In all the presented scenarios, QRD SSEP has much lower complexity than VEP, from -10% for small 4×4 MIMO without auto-iteration to -58% for massive 100×100 MIMO cases with high order constellations and two auto iterations which are the scenarios where VEP might be too complex to be implemented. The second percentage is the complexity of QRD SSEP over QRD SEP to show the additional marginal cost of the proposed algorithm compared to its closest competitor QRD SEP. The two algorithms have very similar complexities per auto iteration, but their performance is vastly different, as shown in the next section. Finally, QRD SSEP SIC has the same computational complexity as QRD SSEP since they compute as many messages for detection (complexity for decoding is omitted here). However, the serial activation of the decoders in the MU case can cause, in certain implementations, a higher decoding latency, which is typical for a SIC detector. Compared to (turbo)-MMSE-SIC [17], [18] or VEP-SIC, this detector does not require to do more operations than its non-SIC variant because of the scalar nature of its underlying FG.

TABLE I
RELATIVE DETECTION COMPLEXITIES: QRD SSEP / VEP - QRD SSEP/QRD SEP

MIMO scenarios	Auto iterations		
	0	1	2
4×4 - 16 QAM	93% - 103%	84% - 103%	81% - 103%
10×10 - 8 PSK	80% - 103%	62% - 104%	56% - 104%
32×32 - 16 QAM	73% - 101%	52% - 101%	46% - 102%
100×100 - 64 QAM	71% - 100%	50% - 100%	42% - 100%

IV. PERFORMANCE RESULTS

Fig. 3 shows the Achievable Information Rate (AIR), which is the maximum data rate in "bits/channel use" that could be achieved with ideal coding. The data rate upper bound is 128 bits/channel use for a 32×32 SU-MIMO with a 16 QAM constellation. The dashed lines represent the AIR of several detectors using BICM (without turbo-iteration), while the solid lines represent the BICM-ID (with turbo-iteration). SEP [5] (resp. QRD SEP [9]) cannot achieve the maximum data rate and is stuck at 50 (resp. 78) bits/channel use with BICM, whereas it can reach 93 (resp. 110) bits/channel use with BICM-ID. QRD PIC BP [13] can achieve the same performance as QRD SEP with BICM and saturate at 102 bits/channel use with BICM-ID. The proposed QRD SSEP can achieve almost the same AIR as VEP [6] and DEP [7] with BICM, but it cannot reach the maximum data rate and is stuck at 120 bits/channel use while VEP and DEP achieve it. The gain of AIR between the proposed algorithm and SEP / QRD SEP is visible starting from code rates $R = 1/3$ and higher for BICM, and from rate $1/2$ and higher for BICM-ID.

Fig. 4 shows the Signal-to-Noise Ratio (SNR) needed to achieve a Frame Error Rate (FER) of 10^{-2} given a computational complexity for each algorithm through turbo-iteration. The SU-MIMO scenarios studied are a 32×32 MIMO with a 16 QAM, shown with "+" markers, and a 10×10 MIMO with an 8 PSK, shown with "o" markers. The SNR is defined as $\frac{E_s}{N_0}$ with $E_s = N_t P_s$ and P_s the energy of one QAM symbol, the constellation power is normalized to 1. 5G compliant LDPC of length $K_b = 3840$ and of rate $1/2$ and $2/3$ are used. The LDPC decoder makes five inner iterations per turbo iteration, and the internal messages of the decoder are kept and reused as initialization messages for the next turbo iteration. A maximum of 9 turbo iterations (i.e. ten decoder calls) brings the overall LDPC iterations to 50. The y-axis measures the detection complexity for a given number of turbo iterations (each point of a curve is associated with a turbo iteration), starting from $T = 0$ with the lowest and least complex point, to $T = 9$ with the highest and most complex point. All detectors use only $L = 1$ auto-iteration per turbo-iteration.

With the 10×10 MIMO scenario with $R = 1/2$, the proposed algorithm, QRD SSEP, has a better performance-to-complexity trade-off than VEP/DEP for a fixed number of turbo-iterations as its performance is similar to VEP/DEP. Compared to SEP and QRD SEP, it has a far better performance-to-complexity trade-off as it can achieve better performance with fewer turbo iterations. When using an LDPC

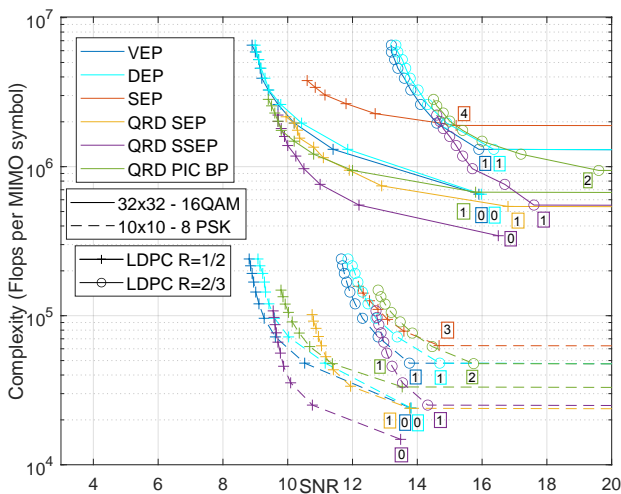


Fig. 4. Performance vs. complexity in flops per MIMO symbol for two ST-BICM-ID SU-MIMO scenarios. Numbers show the first turbo iteration number T per detector, which allows reaching the target FER of 10^{-2} .

code of $R = 2/3$, the same behavior occurs between QRD SSEP and VEP/DEP, while SEP and QRD SEP are not able at all to achieve a FER of 10^{-2} , so there are no curves representing them.

For the 32×32 MIMO with a 16 QAM scenario with $R = 1/2$, QRD SSEP is always less complex than VEP/DEP. It achieves a FER of 10^{-2} with an SNR slightly higher than its competitor. With a few turbo iterations, QRD SSEP performs significantly better than QRD SEP and SEP, and performs slightly better with nine turbo iterations. Like for the previous scenario, with $R = 2/3$, SEP and QRD SEP cannot achieve a FER of 10^{-2} , so there are no curves for them. VEP/DEP can perform better than QRD SSEP, but with a few turbo iteration QRD SSEP performs better for a fixed complexity.

Fig. 5 shows the performance of the algorithms for a BICM-ID 4×8 MU-MIMO with a 64QAM constellation. Only the first, second, and ninth turbo iterations are presented. As in SU-MIMO, SEP performs poorly without many turbo-iterations (Turbo 0 and 1), while QRD SEP performs better. But it is still far from the VEP performance. The proposed SSEP QRD achieves the same performance as VEP for each turbo-iteration, but with a smaller complexity, as shown in Tab. I. The proposed SSEP QRD SIC outperforms VEP with a smaller complexity with a few turbo iterations (Turbo 0 and 1). With nine turbo-iterations, SSEP QRD SIC is still the best-performing MU-MIMO detector, but VEP is very close, as well as SSEP QRD. SEP and SEP QRD are not so far, but they still lack behind by 0.25-0.5dB.

V. CONCLUSION

In this article, we proposed two new algorithms, SSEP QRD for SU-MIMO and MU-MIMO and SSEP QRD SIC for MU-MIMO, that can achieve the same or better performance with respect to VEP and DEP. They both have similar low complexity, lower than VEP/DEP, since they do not require a costly matrix inversion. This makes them great candidates for efficient MIMO and mMIMO detection. This proposal is an additional step toward filling the gap between vector and scalar approaches of EP while keeping a low complexity.

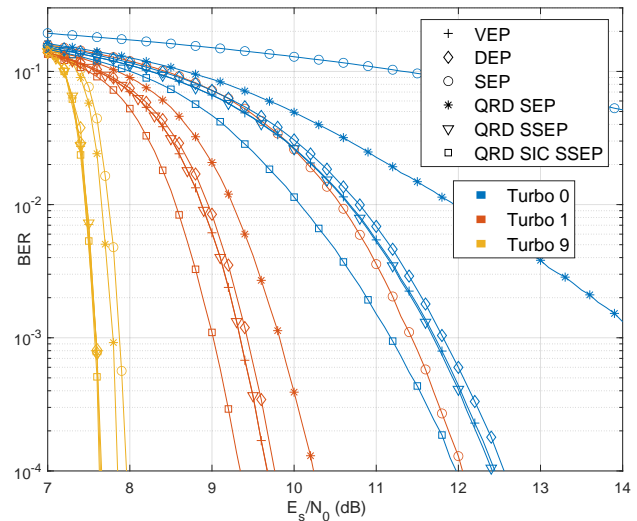


Fig. 5. Performance of BICM-ID MU-MIMO detection with 4×8 MIMO 64QAM on a Rayleigh channel using a LDPC ($K_b = 3840$, $R = 1/2$).

REFERENCES

- [1] T. Lv and F. Long, "Graph-based low complexity detection algorithms in multiple-input-multiple-out systems: an edge selection approach," *IET Commun.*, vol. 7, pp. 1202–1210, 2013.
- [2] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *IEEE Int. Symp. On Inf. Theory*, St. Petersburg, Russia, 2011, pp. 2168–2172.
- [3] P. Schniter *et al.*, "Vector approximate message passing for the generalized linear model," in *50th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, 2016, pp. 1525–1529.
- [4] T. P. Minka, "A Family of Algorithms for Approximate Bayesian Inference," Ph.D. dissertation, MIT, USA, 2001.
- [5] S. Wu *et al.*, "Low-Complexity Iterative Detection for Large-Scale Multiuser MIMO-OFDM Systems Using Approximate Message Passing," *IEEE J. Sel. Top. Signal Process.*, vol. 8, pp. 902–915, 2014.
- [6] M. Senst and G. Ascheid, "How the Framework of Expectation Propagation Yields an Iterative IC-LMMSE MIMO Receiver," in *IEEE Glob. Commun. Conf.*, Houston, TX, USA, 2011, pp. 1–6.
- [7] J. J. Muriillo-Fuentes *et al.*, "A Low-Complexity Double EP-Based Detector for Iterative Detection and Decoding in MIMO," *IEEE Transactions on Communications*, vol. 69, pp. 1538–1547, Mar. 2021.
- [8] S. Park and S. Choi, "QR decomposition aided belief propagation detector for MIMO systems," *Electron. Lett.*, vol. 51, pp. 873–874, 2015.
- [9] Y. Dong *et al.*, "Efficient EP Detectors Based on Channel Sparsification for Massive MIMO Systems," *IEEE Commun. Lett.*, vol. 24, pp. 539–542, 2020.
- [10] X. Meng *et al.*, "Advanced NOMA Receivers from a Unified Variational Inference Perspective," *IEEE J. Select. Areas Commun.*, pp. 1–1, 2021.
- [11] A. Tonello, "Space-time bit-interleaved coded modulation with an iterative decoding strategy," in *IEEE VTS Fall.*, vol. 1, Boston, MA, USA, 2000, pp. 473–478.
- [12] T. P. Minka, "Divergence Measures and Message Passing," Microsoft, TR MSR-TR-2005-173, Jan. 2005.
- [13] A. Mekhiche *et al.*, "Performance-Complexity Trade-Off for Low-Complexity MIMO Detection: simplified BP vs. EP Receivers," in *2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring)*, Helsinki, Finland, 2022, pp. 1–7.
- [14] G. Kallel, "Channel equalization for block transmission systems," *IEEE J. Sel. Areas in Commun.*, vol. 13, pp. 110–121, 1995.
- [15] P. Wolniansky *et al.*, "V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. URSI Int. Symp. Signals Syst. Electron.*, Pisa, Italy, 1998, pp. 295–300.
- [16] M. Senst, "On the Design of Iterative Wireless Receivers: The Divergence Minimization Approach to Approximate Bayesian Inference," Ph.D. dissertation, RWTH Aachen, 2016.
- [17] M. Tüchler and A. C. Singer, "Turbo Equalization: An Overview," *IEEE Trans. Inf. Theory*, vol. 57, pp. 920–952, 2011.
- [18] X. Wang and H. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, pp. 1046–1061, 1999.