Expectation Propagation on Factor Graphs Based on Matrix Decomposition

Adam Mekhiche, Antonio Maria Cipriano, Charly Poulliat

EXPECTATION PROPAGATION ON FACTOR GRAPHS BASED ON MATRIX DECOMPOSITION

Adam Mekhiche[*†]*, Antonio Maria Cipriano*[*]*, Charly Poulliat*[†]

[*] THALES, Gennevilliers, France, Email: name.surname@thalesgroup.com
[†] Univ. of Toulouse, IRIT-INPT, CNRS, Toulouse, France, Email: name.surname@toulouse-inp.fr

ABSTRACT

In the context of the Gaussian linear model, recent works have studied factor graph modification using QR decomposition that enables the derivation of scalar Expectation Propagation (EP) based detectors. In this paper, we investigate on new factor graph representations induced by the use of the Golub-Kahan bi-diagonal Decomposition (GKD) and of the Singular Value Decomposition (SVD). New EP messages induced by the GKD or SVD underlying graphs are derived, that can be both scalar or vector messages. Complexity and performance of the resulting algorithms are studied for digital communications applications.

Index Terms— Factor Graph, Message Passing, Expectation Propagation, Matrix Decomposition, Linear Model

1. INTRODUCTION

In this paper, we consider the Gaussian linear model defined as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \tag{1}$$

where $\mathbf{x} \in \mathbb{C}^{N_t}$, $\mathbf{y} \in \mathbb{C}^{N_r}$, $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$, and $\mathbf{w} \in \mathbb{C}^{N_r}$ is a vector of White Gaussian Noise (WGN) samples with $\mathbb{E}(\mathbf{w}) = 0$ and $\mathbb{E}(\mathbf{w}\mathbf{w}^H) = N_0\mathbf{I}_{N_r}$. This model is widely and extensively used in numerous applications in signal processing. In particular, in digital communications, it has been used for equalization [1], detection [2, 3], channel estimation [4], just to mention a few. It can be represented using a Factor Graph (FG) on which vector or scalar message passing algorithms (MPAs) can be applied to estimate $\mathbf{x}$. Several MPAs could be applied on the FG like Belief Propagation (BP) [4], Approximate Message Passing (AMP) [5], Vector Approximate Message Passing (VAMP) [6] or Generalized Approximate Message Passing (GAMP) [7]. In this work, we focus on MPAs on FG derived within the Expectation Propagation (EP) framework [8], which is another competitive approach to derive efficient MPAs. EP is an approximate Bayesian inference technique that can be interpreted as a generalization of Belief Propagation (BP) within the Message Passing (MP) framework.

By decomposing $\mathbf{H}$, referred to as the channel matrix, and pre-processing the signal $\mathbf{y}$ accordingly, the corresponding FG can be opportunely preprocessed. Several matrix decompositions can be used, like the QR decomposition (QRD), the Golub-Kahan bi-diagonal Decomposition (GKD) [9], all the way to the Singular Value Decomposition (SVD) to alter/modify the underlying graph. Prior works [10, 11] applied QRD and corresponding pre-processing using MPAs. The QRD produces a fairly similar FG compared to its initial form, except fewer messages need to be exchanged but the overall structure remains the same. Other works, e.g. [5, 6], use SVD inside the MPAs in order to efficiently compute matrix inverses. However, they do not apply SVD to pre-process the channel and observations *before* the use of MPA, therefore the underlying graph remains unchanged. Unitary AMP (UAMP) [12] uses a SVD on the linear model but it does not profit from the diagonal matrix sparsity nor it treats the decomposed matrices independently, thus the underlying graph remains the same.

In this paper, we introduce new FG representations related to both GKD and SVD pre-processings, and for which new EP message derivations are obtained. Complexities for the different approaches are then evaluated and performance are illustrated in two main applications, ie. symbol detection in a multiple-input multiple output (MIMO) system [3] and Single Carrier (SC) equalization [13, 14].

The paper is organized as follows. In Section 2, FG representation of the linear model is discussed for different matrix decompositions. Section 3 presents the EP messages derivation induced by the matrix decomposition pre-processing. Computational complexities are discussed in Section 4. Some performance results are given in Section 5 and finally, conclusions and future research perspectives are drawn in Section 6.

2. FACTOR GRAPH REPRESENTATION WITH MATRIX DECOMPOSITION

Factor Graphs (FG) are graphical representations associated with the factorization of a given probability distribution, in our case the a posteriori probability (APP) distribution $P(\mathbf{x}|\mathbf{y}, \mathbf{H})$. The different random variables of the system are associated with *variable nodes* (circles) in the graph, while "constraints" on these variables are represented through *factor* or *function nodes* (squares). The FG representation will depend on the factorization of the APP. Prior partial knowledge of $\mathbf{x}$ and perfect knowledge of $\mathbf{H}$ are assumed in this work as it is often the case with digital communication contexts. Since $\mathbf{w}$ is considered as white, the APP can be factorized for a coded communication system as:

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

where $f_\cdot^{\text{EQU}}$ are likelihood functions, $f_\cdot^{\text{DEM}}$ are the factors associated to the demodulation process and $f_\cdot^{\text{PRI}}$ are priors associated to coded binary vector labels $\mathbf{c}_i$ that map to constellation symbols x_i, $\forall i \in [\![1, N_t]\!]$. This factorization is used in [3, 10, 11] and referred to as scalar factorization. The scalar FG drawn from (2) is shown in Fig. 1. One can notice that, $\forall i, j \in [\![1, N_t]\!], [\![1, N_r]\!]$, the subgraph between function nodes f_j^{EQU} and variable nodes x_i is fully

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

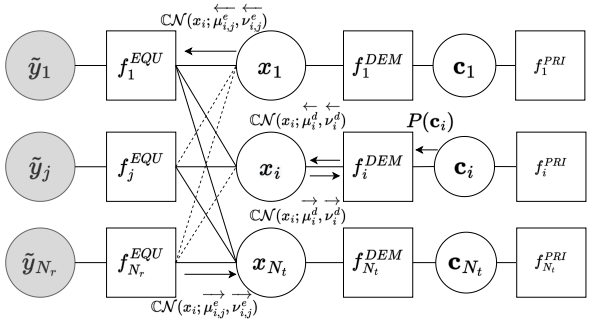

Fig. 1. Factor Graph of $P(\mathbf{x}|\mathbf{y},\mathbf{H})$ factorized as in (2). Dotted edges are removed after QRD pre-processing and variable nodes in grey are known variables.

connected as it expresses the dense nature of the channel matrix. To enable a sparser FG representation (ie. less connected), matrix decomposition techniques can be used, leading to MPAs that can combine both scalar and vector messages depending on the type of the matrix decomposition.

Assuming $\mathbf{H}$ known at the receiver, previous works [10, 11] use QR Decomposition (QRD) [9] to split the channel into two matrices, with $\mathbf{H} = \mathbf{QR}$, where $\mathbf{Q}$ is a unitary matrix and $\mathbf{R}$ is an upper triangular matrix. Then, the observation model can be updated as follows:

$$\tilde{\mathbf{y}} = \mathbf{Q}^H\mathbf{y} = \mathbf{R}\mathbf{x} + \mathbf{Q}^H\mathbf{w} \tag{3}$$

Statistical properties of the additive noise remain unchanged due to the unitary property of $\mathbf{Q}$. The new FG is shown in Fig. 1 with the dotted segments corresponding to removed edges thanks to the QRD.

We now consider two other possible decompositions and analyze the resulting factor graphs. The Golub-Kahan bi-diagonal Decomposition (GKD) [9] of $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ ($N_r \geq N_t$) gives $\mathbf{H} = \mathbf{QBZ}$ with $\mathbf{Q} \in \mathbb{C}^{N_r \times N_r}$ and $\mathbf{Z} \in \mathbb{C}^{N_t \times N_t}$ being unitary matrices while $\mathbf{B} \in \mathbb{C}^{N_r \times N_t}$ is an upper bi-diagonal matrix. The corresponding modified observation model is given by

$$\tilde{\mathbf{y}} = \mathbf{Q}^H\mathbf{y} = \mathbf{BZ}\mathbf{x} + \mathbf{Q}^H\mathbf{w} = \mathbf{B}\bar{\mathbf{x}} + \mathbf{Q}^H\mathbf{w}. \tag{4}$$

where $\bar{\mathbf{x}} = \mathbf{Z}\mathbf{x}$. This leads to the following factorization:

$$P(\mathbf{x}|\mathbf{y},\mathbf{H}) \propto \prod_{j=1}^{N_r} P(\tilde{y}_j|\bar{\mathbf{x}},\mathbf{B}) P(\bar{\mathbf{x}}|\mathbf{x}) \prod_{i=1}^{N_t} P(x_i|\mathbf{c}_i) P(\mathbf{c}_i). \tag{5}$$

Using (5), the new FG is presented in Fig. 2. The graph has a new function node $f^{\bar{\mathbf{x}}|\mathbf{x}}$ which represents the link between $\bar{\mathbf{x}}$ and $\mathbf{x}$. The last decomposition studied in this work is the Singular Value Decomposition (SVD), which can be efficiently calculated starting from GKD. In this latter case, we have $\mathbf{H} = \mathbf{U\Sigma V}$ with $\mathbf{U} \in \mathbb{C}^{N_r \times N_r}$ and $\mathbf{V} \in \mathbb{C}^{N_t \times N_t}$ being unitary matrices, while $\mathbf{\Sigma} \in \mathbb{C}^{N_r \times N_t}$ is a diagonal matrix. Based on the corresponding modified observation model, the factorization (5) remains valid with $\bar{\mathbf{x}} = \mathbf{Vx}$, $\mathbf{Q} = \mathbf{U}$ and $\mathbf{B} = \mathbf{\Sigma}$. It is worth noting that $\forall j \in [\![1, N_r - 1]\!]$, there is no edge between f_j^{EQU} and $\bar{x}_{j+1}$ but only "parallel" edges like on the demapping side of the graph.

Once the graph is appropriately represented, one can derive EP messages along these graphs to estimate the variable nodes x_i. It is worth noting that a vector factorization of $P(\mathbf{x}|\mathbf{y},\mathbf{H})$ like in [2] cannot benefit from those decompositions as the equivalent vector FG is not sensitive to matrix decomposition.

3. EXPECTATION PROPAGATION OVER GKD OR SVD FACTOR GRAPHS

Expectation Propagation (EP) [8] is an approximate Bayesian inference technique that can be interpreted as a generalization of Belief Propagation (BP) within the Message Passing (MP) framework. EP aims at tracking a distribution p by approximating it with a distribution q which lays in a specified set $\mathcal{Q}$ of "simple" distributions. q is the distribution in $\mathcal{Q}$ which is the closest to p according to the inclusive KullBack-Leibler divergence D_{KL}. So the projection of p on $\mathcal{Q} = \mathbb{C}\mathcal{N}(\mu, \nu)$ problem can be written as:

$$\text{proj}_{\mathcal{Q}}[p(x_i)] = q(x_i) = \underset{\tilde{q}\in\mathcal{Q}}{\arg\min}\, D_{KL}(p(x_i)||\tilde{q}(x_i)) \tag{6}$$

The projection is equivalent to moment matching, i.e., the mean and covariance of p and q are the same [8]. The message sent on an edge from a factor node f_j to a variable node x_i of the FG is:

$$m_{f_j\to x_i}(x_i) = \frac{1}{m_{x_i\to f_j}(x_i)} \times \text{proj}_{\mathcal{Q}}\left[m_{x_i\to f_j}(x_i) \int_{\mathbf{x}} f_j(\mathbf{x}) \prod_{i'\in\mathcal{N}(f_j), i'\neq i} m_{x_{i'}\to f_j}(x_{i'}) \right] \tag{7}$$

with $N(f_j)$ the set of the neighbour variable nodes of f_j. This projection constraint only affects few messages in practice because most of the pre-projection distributions of the graph already lie in the Gaussian family. We define the different messages of the graph as the complex Gaussians represented on Fig. 1. Authors from [3] show the application of EP on scalar FG without decomposition and authors from [10] show the application of EP on scalar FG with QRD. There are several possible types of iteration in the graph, the first one between equalization nodes and variable nodes x_i is called *inner* iteration, the second between equalization nodes and demapping nodes is called *auto* iteration, and the last between equalization nodes and prior probabilities nodes is called *turbo* iteration. The comparison between proposed scalar EP (SEP) algorithms in [3] and [10] shows that the QRD pre-processing modifies the graph by reducing the number of edges (the dotted edges of Fig. 1). It also reduces the complexity, as there are fewer exchanged messages, and it can improve performance as the graph is less loopy. This QRD factorization does not require a new factor node as the proposed decompositions (ie. GKD and SVD) do.

The apparition of a new function node $f^{\bar{\mathbf{x}}|\mathbf{x}}$ in Fig. 2 modifies the behaviour of the exchanged EP messages compared to Fig 1. The inner-graph between equalization nodes and variable nodes is even sparser than with QRD.

The function node $f^{\bar{\mathbf{x}}|\mathbf{x}}$ creates four new message types (Fig 2.(b, c, f, g)) and the messages coming from this node to either $\bar{x}_i$ or x_i can be computationally demanding as they require a matrix inversion. Indeed, Fig. 2.(c,g) messages are vector messages, so the proposed EP algorithms combine scalar and vector messages, and are referred to as Hybrid EP (HEP) algorithms. The pre-projection pdf $\tilde{q}(\bar{\mathbf{x}})$ of the message Fig. 2.(c) is a multivariate complex Gaussian of covariance and mean :

$$\bar{\mathbf{V}}^Z = \left(\left(\overrightarrow{\bar{\mathbf{V}}^Z} \right)^{-1} + \left(\mathbf{Z}\overleftarrow{\mathbf{V}^d}\mathbf{Z}^H \right)^{-1} \right)^{-1} \tag{8}$$

$$\bar{\boldsymbol{\mu}}^Z = \bar{\mathbf{V}}^Z \left(\left(\overrightarrow{\bar{\mathbf{V}}^Z} \right)^{-1} \overrightarrow{\bar{\boldsymbol{\mu}}^Z} + \mathbf{Z} \left(\overleftarrow{\mathbf{V}^d} \right)^{-1} \overleftarrow{\boldsymbol{\mu}^d} \right) \tag{9}$$

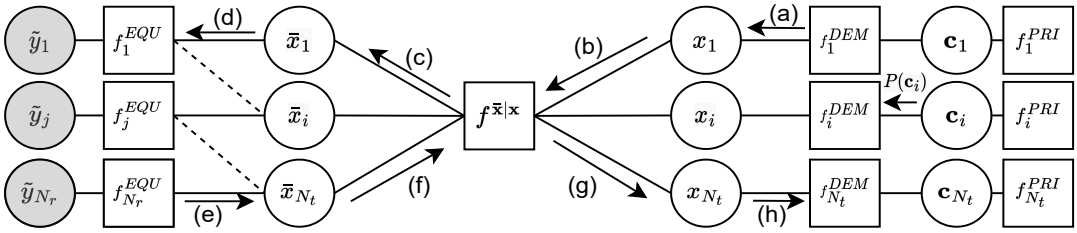

Fig. 2. Factor Graph of $P(\mathbf{x}|\mathbf{y},\mathbf{H})$ factorized as in (5) considering $N_t = N_r$. Dotted edges are removed after SVD and greyed variable nodes are known variables. (a): $\mathbb{CN}\left(\overleftarrow{\mu_i^d},\overleftarrow{\nu_i^d}\right)$, (b): $\mathbb{CN}\left(\overleftarrow{\mu_i^z},\overleftarrow{\nu_i^z}\right)$, (c): $\mathbb{CN}\left(\overleftarrow{\bar{\mu}_i^z},\overleftarrow{\bar{\nu}_i^z}\right)$, (d): $\mathbb{CN}\left(\overleftarrow{\bar{\mu}_{i,j}^e},\overleftarrow{\bar{\nu}_{i,j}^e}\right)$, (e): $\mathbb{CN}\left(\overrightarrow{\bar{\mu}_{i,j}^e},\overrightarrow{\bar{\nu}_{i,j}^e}\right)$, (f): $\mathbb{CN}\left(\overrightarrow{\bar{\mu}_i^z},\overrightarrow{\bar{\nu}_i^z}\right)$, (g): $\mathbb{CN}\left(\overrightarrow{\mu_i^z},\overrightarrow{\nu_i^z}\right)$, (h): $\mathbb{CN}\left(\overrightarrow{\mu_i^d},\overrightarrow{\nu_i^d}\right)$.

with $\overleftarrow{\mathbf{V}^d} = \text{diag}(\overleftarrow{\nu_1^d},\cdots,\overleftarrow{\nu_{N_t}^d})$, $\overrightarrow{\mathbf{V}^Z} = \text{diag}(\overrightarrow{\nu_1^Z},\cdots,\overrightarrow{\nu_{N_t}^Z})$, $\overleftarrow{\boldsymbol{\mu}^d} = [\overleftarrow{\mu_1^d},\cdots,\overleftarrow{\mu_{N_t}^d}]^T$ and $\overrightarrow{\bar{\boldsymbol{\mu}}^Z} = [\overrightarrow{\bar{\mu}_1^Z},\cdots,\overrightarrow{\bar{\mu}_{N_t}^Z}]^T$. Unlike the EP messages, the a posteriori pdf $\tilde{q}(\tilde{\mathbf{x}})$ does not have a diagonal covariance matrix so the matrix inversion is computationally demanding. The same behavior occurs for Fig. 2.(g) posterior $\tilde{q}(\mathbf{x})$, which has a covariance matrix and mean vector:

$$\mathbf{V}^Z = \left(\left(\overleftarrow{\mathbf{V}^d}\right)^{-1} + \left(\mathbf{Z}\overrightarrow{\mathbf{V}^Z}\mathbf{Z}^H\right)^{-1}\right)^{-1} \tag{10}$$

$$\boldsymbol{\mu}^Z = \mathbf{V}^Z\left(\mathbf{Z}^H\left(\overrightarrow{\mathbf{V}^Z}\right)^{-1}\overrightarrow{\bar{\boldsymbol{\mu}}^Z} + \left(\overleftarrow{\mathbf{V}^d}\right)^{-1}\overleftarrow{\boldsymbol{\mu}^d}\right) \tag{11}$$

The computational complexity issue is tackled naturally as many scenarios. For instance the first auto-iteration has the particularity that $\overrightarrow{V^z}$ and $\overleftarrow{V^z}$ are scalar diagonal matrices which translates into, $\forall i \in [\![1,N_t]\!]$, $m_{f^{\bar{\mathbf{x}}|\mathbf{x}}\to x_i}(x_i) = \mathbb{CN}(x_i;\sum_{i'=1}^{N_t} z_{i,i'}\overrightarrow{\bar{\mu}_i^z},\overrightarrow{\bar{\nu}_i^z})$. For the return message Fig. 2.(g), the posterior variance can be computed by inverting a diagonal matrix since $\overleftarrow{\mathbf{V}^Z}$ is a scalar diagonal matrix. To auto-iterate while keeping the complexity low, the message Fig. 2.(f) variance can be set to a scalar diagonal matrix using the minimum of variances which makes the inversion a diagonal matrix inversion without losing any performance during the Fig. 2.(c) message update. The choice of using the minimum instead of the mean of variances is an heuristic from the studied cases of Section 5.

The study of $\mathbf{Z}$ (for GKD pre-processing) shows that its first row and first column are vectors with a one as first element and zeros for the rest which highlights the fact that the function node between x_1 and $\bar{x}_1$ could be represented as an identity function. The matrix inversion is less complex as we can work with a smaller matrix of size $(N_t-1)\times(N_t-1)$. The Hybrid EP with Golub-Kahan decomposition (HEP GKD) algorithm is shown in Alg. 1. Note that, in Alg. 1, non informative Gaussians refer to as complex Gaussian distributions with a zero mean and infinite variance. Furthermore, Alg. 1 can take soft input so it could also be used in a BICM-ID (Bit Interleaved Coded Modulation with Iterative Decoding) scheme. The only differences in the messages of HEP SVD compared to HEP GKD is the computation of the messages Fig. 2.(d,f). In HEP SVD, $\forall i \in [\![1,N_t]\!]$, $m_{\bar{x}_i\to f^{\bar{\mathbf{x}}|\mathbf{x}}}(\bar{x}_i) = m_{f_i^{\text{EQU}}\to\bar{x}_i}(\bar{x}_i)$ and $m_{\bar{x}_i\to f_i^{\text{EQU}}}(\bar{x}_i) = m_{f^{\bar{\mathbf{x}}|\mathbf{x}}\to\bar{x}_i}(\bar{x}_i)$ which are simpler to compute.

4. COMPUTATIONAL COMPLEXITY

In this section, floating point operations (flops) are used to quantify the computational complexity of the algorithms. According to [9], the QRD complexity order is $\mathcal{O}(N_t^3/3)$ flops. The GKD can be efficiently computed starting with a QRD, and its overall complexity is $\mathcal{O}(2N_rN_t^2+2N_t^3)$ flops. Finally, SVD is efficiently computed starting with GKD, and its total complexity is in the order of $\mathcal{O}(2N_rN_t^2+11N_t^3)$ flops. Note that these pre-processings need to be executed only when the channel changes, e.g. in a block fading scenario only once per block. Those complexities are to be compared to the full-rank matrix inversion, required in LMMSE or Vector EP [2], complexity which is $\mathcal{O}(N_t^3)$. Thanks to the proposed simplification which removes the need of inverting a matrix to compute the message Fig. 2.(c), the amount of matrix inverses in HEP GKD and HEP SVD is lower than in VEP if there are auto or turbo iterations.

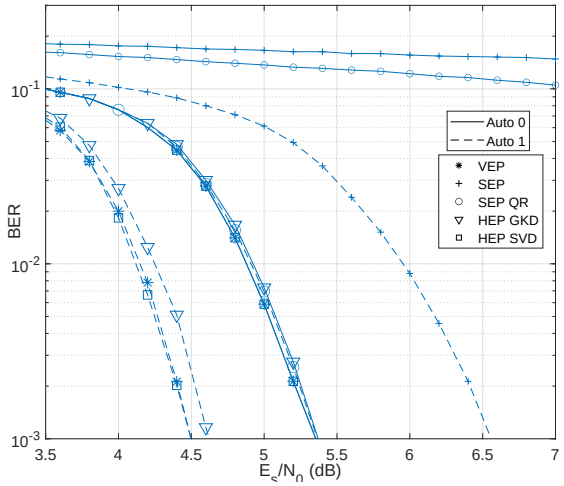

Fig. 3. Detection performances of 32×32 coded MIMO-OFDM

5. NUMERICAL RESULTS

To assess the performance of the proposed algorithms, we first consider a BICM MIMO-OFDM (orthogonal frequency division multiplexing) detection scheme. We use a 5G LDPC (low-density parity-check) error correcting code without turbo-iteration, of length $N = 2048$ and coding rate $r = 1/2$ using the BP algorithm with 50 iterations. The resulting codewords are then mapped to QPSK symbols after random interleaving. The symbols are sent via $N_t = 32$ transmit antennas through an ergodic flat Rayleigh channel using a $N_r = 32$ receiving antennas receiver. The Bit Error Rate (BER) versus the energy-per-symbol to noise ratio E_s/N_0 is reported in Fig. 3,

Algorithm 1 Hybrid EP with GKD

Input: $\mathbf{y}, \mathbf{H}, N_0, \lambda^a$ (prior Log Likelihood Ratio - LLR)

Output: $P(\mathbf{x}|\mathbf{y},\mathbf{H}) \sim \mathbb{CN}\left(\mathbf{x}; \overrightarrow{\boldsymbol{\mu}^d}, \overrightarrow{\mathbf{V}^d}\right)$

$\mathbf{H} = \mathbf{QBZ}, \quad \mathbf{y} \leftarrow \mathbf{Q}^H\mathbf{y}, \quad \mathbf{H} \leftarrow \mathbf{B}, \bar{\mathbf{x}} = \mathbf{Z}\mathbf{x}$

Set Fig. 2.(e).(f).(g).(h) to non informative gaussians

for $a = 0$: auto-iterations **do**

Compute Fig. 2. (a) and Fig. 2. (b):

for $i = 1 : N_t$ **do**

$\tilde{q}(x_i) \propto \exp\left(-\frac{|x_i - \overrightarrow{\mu_i^d}|^2}{\overrightarrow{\nu_i^d}} - \sum_{n=1}^{N} c_{i,n}\lambda_{i,n}^a\right)$

$q(x_i) \sim \mathbb{CN}\left(x_i; \mu_i^d = \mathbb{E}\left[\tilde{q}(x_i)\right], \nu_i^d = \mathrm{Var}\left[\tilde{q}(x_i)\right]\right)$

$\overleftarrow{\mu_i^d} = \overleftarrow{\nu_i^d}\left(\frac{\mu_i^d}{\nu_i^d} - \frac{\overrightarrow{\mu_i^d}}{\overrightarrow{\nu_i^d}}\right)$ and $\overleftarrow{\nu_i^d} = \left(\frac{1}{\nu_i^d} - \frac{1}{\overrightarrow{\nu_i^d}}\right)^{-1}$

$\overleftarrow{\mu_i^z} = \overleftarrow{\mu_i^d}$ and $\overleftarrow{\nu_i^z} = \overleftarrow{\nu_i^d}$

end for

Compute Fig. 2. (c):

Pre-projection $\tilde{q}(\bar{\mathbf{x}})$ computed as in (9)

for $i = 1 : N_t$ **do**

$\overleftarrow{\nu_i^Z} = \left(\frac{1}{\nu_i^Z} - \frac{1}{\overrightarrow{\nu_i^Z}}\right)^{-1}$ and $\overleftarrow{\mu_i^Z} = \overleftarrow{\nu_i^Z}\left(\frac{\mu_i^Z}{\nu_i^Z} - \frac{\overrightarrow{\mu_i^Z}}{\overrightarrow{\nu_i^Z}}\right)$

end for

Compute Fig. 2. (d):

for $i = 1 : N_t$ and $j = 1 : N_r$ **do**

$\overleftarrow{\nu_{i,j}^e} = \left(\frac{1}{\overleftarrow{\nu_i^Z}} - \frac{1}{\overrightarrow{\nu_{i,j'}^e}}\right)^{-1}, \overleftarrow{\mu_{i,j}^e} = \overleftarrow{\nu_{i,j}^e}\left(\frac{\overleftarrow{\mu_i^Z}}{\overleftarrow{\nu_i^Z}} - \frac{\overrightarrow{\mu_{i,j'}^e}}{\overrightarrow{\nu_{i,j'}^e}}\right)$ with $j' \neq j$ and $j' \in \mathcal{N}(\bar{x}_i)$

end for

Compute Fig. 2. (e).(f):

for $j = 1 : N_r$ and $i = 1 : N_t$ **do**

$\sigma_{i,j}^2 = N_0 + \sum_{i' \neq i} |h_{j,i'}|^2 \overleftarrow{\nu_{i',j}^e}, z_{i,j} = \tilde{y}_j - \sum_{i' \neq i} h_{j,i'} \overleftarrow{\mu_{i',j}^e}$

$\overrightarrow{\mu_{i,j}^e} = \frac{z_{i,j}}{h_{j,i}}$ and $\overrightarrow{\nu_{i,j}^e} = \frac{\sigma_{i,j}^2}{|h_{j,i}|^2}$

$\overrightarrow{\nu_i^Z} = \left(\sum_{j' \in \mathcal{N}(x_i)} \frac{|h_{j',i}|^2}{\sigma_{i,j}^2}\right)^{-1}, \overrightarrow{\mu_i^Z} = \overrightarrow{\nu_i^Z} \sum_{j' \in \mathcal{N}(x_i)} \frac{h_{j'i}^* z_{i,j}}{\sigma_{i,j}^2}$

end for

Compute Fig. 2. (g).(h):

Pre-projection $\tilde{q}(\mathbf{x})$ computed as in (11)

for $i = 1 : N_t$ **do**

$\overrightarrow{\nu_i^Z} = \left(\frac{1}{\nu_i^Z} - \frac{1}{\overleftarrow{\nu_i^Z}}\right)^{-1}$ and $\overrightarrow{\mu_i^Z} = \overrightarrow{\nu_i^Z}\left(\frac{\mu_i^Z}{\nu_i^Z} - \frac{\overleftarrow{\mu_i^Z}}{\overleftarrow{\nu_i^Z}}\right)$

$\overrightarrow{\mu_i^d} = \overrightarrow{\mu_i^z}, \overrightarrow{\nu_i^d} = \overrightarrow{\nu_i^z}$

end for // **end for**

with E_s the sum of the energies of the symbols of all antennas. The inner-iteration number of HEP GKD is set to one as the messages on the inner graph, the graph between equalization nodes and variable nodes, are smartly computed from bottom to top (only $2N_t - 1 = 63$ messages compared to the $N_t(N_t+1)/2 = 528$ of SEP QRD and $N_t^2 = 1024$ of SEP per inner-iteration) to achieve near optimal performance. HEP SVD does not perform inner-iteration because it already achieves optimal performance as the symbols $\bar{x}_i$ are independent after SVD pre-processing. The two proposed algorithms, HEP GKD and HEP SVD, without any auto-iteration, have the same performance of the Linear Minimum Mean Square Error (LMMSE) detector [15]. For 1 auto-iteration, HEP SVD has the same performance as VEP [2], while HEP GKD is $0.1 - 0.2$ dB behind for a target BER of 10^{-3}. The classic SEP [3] does not converge without auto-iterations and is still performing worse with more than 1 dB loss at a target BER of 10^{-3}. SEP QRD [10] requires at least one auto-iteration to achieve LMMSE performance while not being able to be close to LMMSE without any auto-iteration. These behaviours can be explained by the dense (fully connected) nature of the inner graph of SEP and the lack of sparsity of the inner graph for SEP QRD.

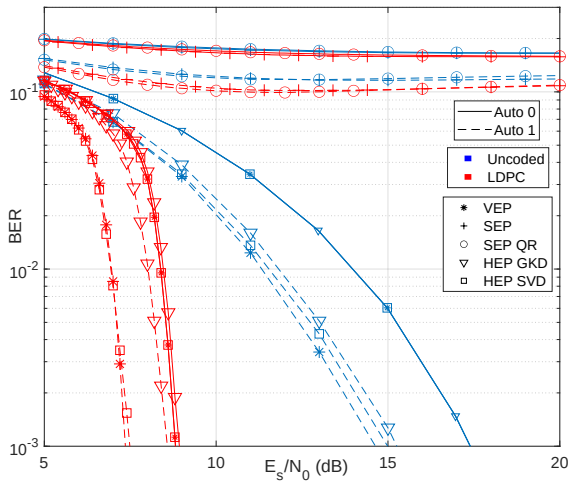

Fig. 4. Performance of a SISO-SC transmission of block 16×18.

Some results for a Single-Carrier (SC) BICM scheme with soft output equalization are also reported in Fig. 4, using a power normalized frequency selective deterministic channel with 3 equal taps at a delay of 0, 1 and 2 symbol times. In our example, $\mathbf{H}$ results in a $N_t = 16$ and $N_r = 18$ channel matrix with the same value on the diagonal and on the following two lower diagonals. A 5G LDPC error correcting code of length $N = 2048$ and $r = 2/3$, using the BP algorithm with 50 iterations, is used without turbo-iteration. HEP GKD and HEP SVD have the same number of inner-iterations as before. For this scenario, SEP and SEP QRD are not able to equalize the signal with or without auto-iteration, contrary to the MIMO-OFDM case. Similarly to the MIMO-OFDM context, coded HEP SVD can achieve the same performance as LMMSE without auto-iteration and as VEP with one auto-iteration. Performance slightly degrades with increasing E_s/N_0. Uncoded HEP SVD performs as well as LMMSE without auto-iteration and as VEP with one auto-iteration. Uncoded HEP GKD achieves LMMSE performance without auto-iteration and achieves almost the same performance as SEP. Coded HEP GKD can almost achieve LMMSE performance without auto-iteration and it performs slightly better with one.

6. CONCLUSION

In this work, we investigate on new factor graph representations for the Gaussian linear model, thanks to matrix decompositions. For these graph representations, we derived two new EP based algorithms. The proposed HEP GKD and HEP SVD benefit from the new graphical representations to enhance the inference accuracy over the SEP QRD algorithm at the cost of a higher complexity which can be reduced with the proposed heuristics. The proposed HEP SVD achieves same performance as VEP with similar complexity while HEP GKD can achieve almost the same performance with a slight reduction of complexity. Performance of the proposed algorithms has been assessed for two main applications in digital communications, i.e. SC equalization and MIMO-OFDM detection, to highlight the benefits of these new techniques compared to state-of-the art algorithms. The proposed FG can be the support on which many other MPAs could be developed, like AMP or BP.

7. REFERENCES

[1] Serdar Sahin, Antonio Maria Cipriano, Charly Poulliat, and Marie-Laure Boucheret, "A Framework for Iterative Frequency Domain EP-Based Receiver Design," *IEEE Transactions on Communications*, vol. 66, no. 12, pp. 6478–6493, Dec. 2018.

[2] Martin Senst and Gerd Ascheid, "How the Framework of Expectation Propagation Yields an Iterative IC-LMMSE MIMO Receiver," in *IEEE Global Telecommunications Conference - GLOBECOM*, Houston, TX, USA, Dec. 2011, pp. 1–6, IEEE.

[3] Sheng Wu, Linling Kuang, Zuyao Ni, Jianhua Lu, Defeng Huang, and Qinghua Guo, "Low-Complexity Iterative Detection for Large-Scale Multiuser MIMO-OFDM Systems Using Approximate Message Passing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 902–915, Oct. 2014.

[4] Ananthanarayanan Chockalingam and Balaji Sundar Rajan, *Large MIMO systems*, Cambridge University Press, New York, 2014.

[5] Xiangming Meng, Lei Zhang, Chao Wang, Lei Wang, Yiqun Wu, Yan Chen, and Wenjin Wang, "Advanced NOMA Receivers from a Unified Variational Inference Perspective," *IEEE Journal on Selected Areas in Communications*, pp. 1–1, 2021.

[6] Philip Schniter, Sundeep Rangan, and Alyson K. Fletcher, "Vector approximate message passing for the generalized linear model," in *50th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, Nov. 2016, pp. 1525–1529, IEEE.

[7] Sundeep Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *IEEE International Symposium on Information Theory Proceedings*, July 2011, pp. 2168–2172, ISSN: 2157-8117.

[8] Thomas P. Minka, *A Family of Algorithms for Approximate Bayesian Inference*, Ph.D. thesis, Massachusetts Institute of Technology (MIT), USA, 2001.

[9] Gene H. Golub and Charles F. Van Loan, *Matrix Computations 4 edition.*, Johns Hopkins University Press, Baltimore, fourth edition edition, 2013.

[10] Yuanyuan Dong, Hua Li, Zhenyu Zhang, Xiyuan Wang, and Xiaoming Dai, "Efficient EP Detectors Based on Channel Sparsification for Massive MIMO Systems," *IEEE Communications Letters*, vol. 24, no. 3, pp. 539–542, Mar. 2020.

[11] Sangjoon Park and Sooyong Choi, "QR decomposition aided belief propagation detector for MIMO systems," *Electronics Letters*, vol. 51, no. 11, pp. 873–874, May 2015.

[12] Man Luo, Qinghua Guo, Ming Jin, Yonina C. Eldar, Defeng Huang, and Xiangming Meng, "Unitary Approximate Message Passing for Sparse Bayesian Learning," *IEEE Transactions on Signal Processing*, vol. 69, pp. 6023–6039, 2021.

[13] Serdar Şahin, Antonio M. Cipriano, Charly Poulliat, and Marie-Laure Boucheret, "Iterative Equalization Based on Expectation Propagation: A Frequency Domain Approach," in *26th European Signal Processing Conference (EUSIPCO)*, Sept. 2018, pp. 932–936, ISSN: 2076-1465.

[14] Irene Santos, Juan Jose Murillo-Fuentes, Rafael Boloix-Tortosa, Eva Arias-de Reyna, and Pablo M. Olmos, "Expectation Propagation as Turbo Equalizer in ISI Channels," *IEEE Transactions on Communications*, pp. 1–1, 2016.

[15] Steven M. Kay, *Fundamentals of statistical signal processing: Estimation theory*, Prentice Hall signal processing series. Prentice-Hall PTR, Englewood Cliffs, N.J, 1993.