Performance-Complexity Trade-Off for Low-Complexity MIMO Detection: simplified BP vs. EP Receivers

Adam Mekhiche, Antonio Maria Cipriano, Charly Poulliat

Performance-Complexity Trade-Off for Low-Complexity MIMO Detection: simplified BP vs. EP Receivers

Adam Mekhiche
Thales, INP Toulouse - ENSEEIHT
Gennevilliers, France
adam.mekhiche@thalesgroup.com

Antonio Maria Cipriano
Thales
Gennevilliers, France
antonio.cipriano@thalesgroup.com

Charly Poulliat
INP Toulouse - ENSEEIHT
Toulouse, France
charly.poulliat@enseeiht.fr

***Abstract*—In this study, we intend to make an in-depth investigation of the performance-complexity trade-off of low complexity Multiple-Input Multiple-Output (MIMO) signal detection based on message passing algorithms. Several detection algorithms such as Belief Propagation (BP) and Expectation Propagation (EP) have been proposed to approximate symbol Maximum A Posteriori (MAP) for high dimensional signaling. We propose a thorough examination of those algorithms and some of their low-complexity versions, through a complexity/performance trade-off analysis to identify modes of operation depending on the number of antennas and constellation order. Finally, we propose a new simplified BP detection scheme, which combines the advantages of QR precoding and Interference Cancellation (IC).**

***Index Terms*—Multiple-Input Multiple-Output, Maximum A Posteriori, Belief Propagation, Expectation Propagation, QR factorisation, Parallel Interference Cancellation**

I. Introduction

Multiple-Input Multiple-Output (MIMO) has become a widely used technology, at the heart of major telecommunication standards like Long Term Evolution (LTE), New radio (NR) and Wi-Fi. The two main improvements of MIMO systems over Single-Input Single-Output (SISO) systems are fading robustness and spectral efficiency thanks to higher spatial diversity and stream multiplexing, respectively, but these gains come with a computational cost [1]. The optimal detection becomes exponentially computationally complex with the number of antennas and polynomial with the modulation order. For MIMO systems of size greater than 4×4 or using high modulation order, the optimal soft detection algorithm, the Maximum A Posteriori (MAP), is hardly considered in practice due to the huge number of possible received symbols. This computational complexity, even greater for massive MIMO (mMIMO), has led to emerging solutions based on Message Passing Algorithms (MPAs), like Expectation Propagation (EP), Belief Propagation (BP) and various variations on Approximate Message Passing (AMP). Low complexity versions of these detection algorithms present a reduced complexity and quasi-optimal performance obtained through iterations within the detector and possibly with the decoder [2][3][4].
While such simplified MPAs are the only viable choices in mMIMO systems, this study aims at having a deeper investigation into their relevance for 5G's "V2V-NR" applications [5]. For vehicular 5G communications, 3GPP has standardized only 2×2 MIMO schemes, and constellations up to 64QAM. Future evolutions will add few antennas, but it is highly probable that V2V will still use small MIMO configurations. It is hence interesting to compare simplified MPAs with the optimal algorithm on complexity and performance basis in this V2V-NR context.

Our investigation focuses on a simplified BP scheme: Parallel Interference Cancellation (PIC) BP, which is compared to EP and to optimal BP and MAP according to a complexity/performance trade-off. Our first contribution is the presentation of a new simplified BP algorithm that benefits both from QR and PIC processing. Our second contribution consists in a fair and precise estimation of the computational complexity. Contrarily to many studies which computes only asymptotic complexity orders, here all operations are taken into account by using realistic weights in FLoating Point OPerations (FLOPs) for each one of them.

The paper is organized as follows: the system model is introduced in Section II. In Section III, EP is presented, followed by BP and its reduced complexity version PIC BP. A new reduced-complexity BP-based algorithm, referred to as QR PIC BP, is introduced too. The methodology for assessing algorithms' computational complexity is proposed in Section IV. It is used to calculate the performance/complexity trade-off. The Symmetric Information Rate (SIR), a theoretical rate upper bound, is discussed to further understand the behavior of such receivers. Finally, conclusions are given and future research perspectives are drawn in Section V.

II. System Model

We consider a $N_t \times N_r$ MIMO system using bit-interleaved coded modulation (BICM) and spatial multiplexing. Let $\mathbb{F}_2$ denotes the binary field. K information bits $\mathbf{b} = [b_1, \cdots, b_K] \in \mathbb{F}_2^K$ are encoded using a binary error correcting code of rate $R = K/N$, here either a convolutional or a Low-Density

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

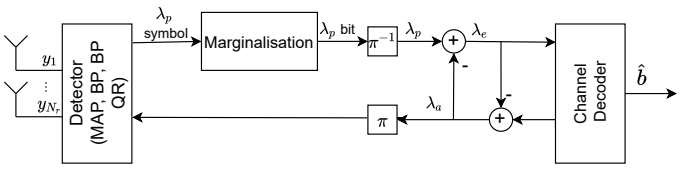

Fig. 1. Block diagram of a generic turbo MIMO receiver with the random interleaving (π) / de-interleaving (π^{-1}) functions.

Parity-Check (LDPC) code. The coded bits $\mathbf{c} = [c_1, \cdots, c_N]$ are interleaved and then Gray-mapped to a constellation $\mathcal{S}$ with $|\mathcal{S}| = M$, each symbol carrying a group of $m = \log_2(M)$ bits. We note $\varphi : \mathbb{F}_2^m \rightarrow \mathcal{S}$ the mapping function and $\varphi_k^{-1} : \mathcal{S} \rightarrow \mathbb{F}_2$, for $k = 1, \ldots, m$, the inverse modulation functions which give the value of the k-th bit of $s \in \mathcal{S}$. The obtained symbols are split across the N_t emitting antennas. Let $\mathbf{s} = [s_1, \cdots, s_{N_t}]^T \in \Psi \subset \mathbb{C}^{N_T}$, where $\Psi \triangleq \mathcal{S}^{N_t}$. The received signal can be written as

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

where $\mathbf{y} = [y_1, \cdots, y_{N_r}]^T$ is the vector of observations, $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ is the channel matrix, whose entries $h_{j,i} \sim \mathcal{CN}(0, 1)$ are supposed to be perfectly known at the receiver $\forall i, j \in [\![1, N_t]\!] \times [\![1, N_r]\!]$. Moreover, $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, N_0\mathbf{I}_{N_r})$, ie. $\mathbf{w} = [w_1, \cdots, w_{N_r}]^T$ is an additive white Gaussian noise vector with mean $\mathbb{E}(\mathbf{w}) = \mathbf{0}$ and covariance matrix $\mathbb{E}(\mathbf{w}\mathbf{w}^H) = N_0\mathbf{I}_{N_r}$, where $\mathbb{E}(.)$, $(.)^H$ and $(.)^T$ denote respectively the expectation operator, the conjugate transpose and transpose operators, and $\mathbf{I}_N$ is the identity matrix of size N. Furthermore, $\forall \eta \in \mathcal{S}$, we define $\Psi_i^\eta \triangleq \{\mathbf{z} \in \Psi | z_i = \eta\}$ and $\forall k \in [\![1 \ldots N_r]\!]$, $\Psi_{i,k}^+ \triangleq \{\mathbf{s} \in \Psi | s_{i,k} = 0\}$ (resp. $\Psi_{i,k}^- \triangleq \{\mathbf{s} \in \Psi | s_{i,k} = 1\}$) with s_i being the i-th symbol of the vector $\mathbf{s}$ and $s_{i,k}$ the k-th bit of the i-th symbol of the vector $\mathbf{s}$ respectively.

Optimal soft output detection can be achieved by applying the MAP algorithm [6], which computes a posteriori log likelihood ratio (LLR) as follows. Let us denote $\lambda_p^{i,k}$ and $\lambda_a^{i,k}, \forall i, k \in [\![1, N_t]\!] \times [\![1, m]\!]$ respectively the a posteriori and a priori LLRs associated with the k-th bit of the symbol sent on the i-th antenna. The a posteriori LLRs is given by

$$\begin{aligned}\lambda_p^{i,k} &= \log\left(\frac{\mathbb{P}(z_{i,k} = 0|\mathbf{y}, \mathbf{H})}{\mathbb{P}(z_{i,k} = 1|\mathbf{y}, \mathbf{H})}\right) \\ &= \log\left(\frac{\sum\limits_{\mathbf{z} \in \Psi_{i,k}^+} e^{-\frac{||\mathbf{y}-\mathbf{H}\mathbf{z}||^2}{N_0}} \prod\limits_{i'}^{N_t} \prod\limits_{k'}^{m} \exp(-z_{i',k'}\lambda_a^{i',k'})}{\sum\limits_{\mathbf{z} \in \Psi_{i,k}^-} e^{-\frac{||\mathbf{y}-\mathbf{H}\mathbf{z}||^2}{N_0}} \prod\limits_{i'}^{N_t} \prod\limits_{k'}^{m} \exp(-z_{i',k'}\lambda_a^{i',k'})}\right)\end{aligned} \tag{2}$$

where the prior LLR $\lambda_a^{i,k} = \log(\mathbb{P}(z_{i,k} = 0)/\mathbb{P}(z_{i,k} = 1))$ is provided by the channel decoder if turbo-detection is used. We also note $\lambda_e^{i,k} = \lambda_p^{i,k} - \lambda_a^{i,k}, \forall i, k \in [\![1, N_t]\!] \times [\![1, m]\!]$ the extrinsic LLR values in case of turbo-detection. Indexing related to the number of turbo iterations is voluntarily omitted for the sake of readability. MAP complexity quickly becomes unbearable, as it scales as $\mathcal{O}(M^{N_t})$. Further simplifications can be done by considering the max-log approximation, where the log-sum-exp in (2) are replaced by max-operations, but its complexity always scales as $\mathcal{O}(M^{N_t})$. The structure of the considered receiver is highlighted in Fig.1.

III. Message Passing Receivers

In this section, we review some efficient low-complexity receivers derived within a message passing framework.

A. Expectation Propagation

Expectation Propagation (EP) is a soft-input soft-output (SISO) message passing approach to solve high computational complexity detection problems. As depicted in Fig. 2, EP exchanges messages between the detection node and the demapping node. The detection node applies an IC-LMMSE equalizer/detector [7], i.e. a linear minimum mean square error spatial filtering with an interference cancellation step. On the demapping node side, two types of messages are computed: one type for the channel decoder which computes the extrinsic LLRs and another one for the detection node which corresponds to the computing of prior probabilities used by the IC-LMMSE equalizer. The inner iterative exchanges between the detector and the demapper are referred to as auto-iterations, while the possible outer iterative exchanges between the demapper and the channel decoder are referred to as turbo iterations. The inner messages, denoted $m_{\text{equ}\rightarrow\text{dem}}$

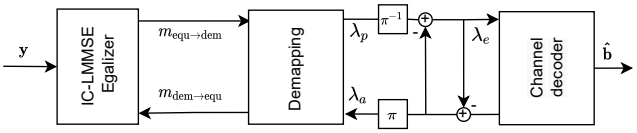

Fig. 2. Block diagram of the EP receiver with the several messages exchanged and the random interleaving (π) / de-interleaving (π^{-1}) functions.

and $m_{\text{dem}\rightarrow\text{equ}}$, are modeled as follow

$$m_{\text{equ}\rightarrow\text{dem}} \propto \mathcal{CN}(\overrightarrow{\boldsymbol{\mu}}, \text{diag}\left\{\overrightarrow{\sigma}_1^2, \ldots, \overrightarrow{\sigma}_{N_t}^2\right\}) \tag{3}$$

$$m_{\text{dem}\rightarrow\text{equ}} \propto C\mathcal{N}(\overleftarrow{\boldsymbol{\mu}}, \text{diag}\left\{\overleftarrow{\sigma}_1^2, \ldots, \overleftarrow{\sigma}_{N_t}^2\right\}) \tag{4}$$

where $\overrightarrow{\boldsymbol{\mu}} = [\overrightarrow{\mu}_1, \cdots, \overrightarrow{\mu_{N_t}}]^T$, $\overleftarrow{\boldsymbol{\mu}} = [\overleftarrow{\mu_1}, \cdots, \overleftarrow{\mu_{N_t}}]^T$, and $\text{diag}\{x_1, \cdots, x_{N_t}\}$ is a diagonal matrix of size $N_t \times N_t$ whose non-zero elements are given by $\{x_1, \cdots, x_{N_t}\}$.
The message $m_{\text{equ}\rightarrow\text{dem}}$ for antenna $i \in [1, \cdots, N_t]$, can be computed as

$$\begin{aligned}\overrightarrow{\mu_i} &= \overleftarrow{\mu_i} + \frac{\mathbf{h}_i^H \mathbf{C}_\mathbf{y}^{-1}\left(\mathbf{y} - \mathbf{H}\overleftarrow{\boldsymbol{\mu}}\right)}{\mathbf{h}_i^H \mathbf{C}_\mathbf{y}^{-1}\mathbf{h}_i} \\ \overrightarrow{\sigma_i}^2 &= (\mathbf{h}_i^H \mathbf{C}_\mathbf{y}^{-1}\mathbf{h}_i)^{-1} - \overleftarrow{\sigma_i}^2\end{aligned} \tag{5}$$

with $\mathbf{h}_i$ the i-th column of the matrix $\mathbf{H}$, and

$$\mathbf{C}_\mathbf{y} = \mathbf{H}\mathbf{C}_\mathbf{x}\mathbf{H}^H + N_0\mathbf{I}_{N_r}, \mathbf{C}_\mathbf{x} = \text{diag}\left\{\overleftarrow{\sigma}_1^2, \ldots, \overleftarrow{\sigma}_{N_t}^2\right\}. \tag{6}$$

The message $m_{\text{dem}\rightarrow\text{equ}}$ is computed, for antenna $i \in [1, \cdots, N_t]$, as

$$\overleftarrow{\mu}_i = \frac{\mu_i \overrightarrow{\sigma}_i^2 - \overrightarrow{\mu_i}\sigma_i^2}{\overrightarrow{\sigma}_i^2 - \sigma_i^2} \quad \text{and} \quad \overleftarrow{\sigma}_i^2 = \frac{\overrightarrow{\sigma}_i^2\sigma_i^2}{\overrightarrow{\sigma}_i^2 - \sigma_i^2} \tag{7}$$

with

$$\mu_i = \sum_{s\in\mathcal{S}} f_i(s)s \quad \text{and} \quad \sigma_i^2 = \sum_{s\in\mathcal{S}} f_i(s)||s-\mu_i||^2. \tag{8}$$

We note $f_i(s)$ the a posteriori categorical distribution

$$f_i(s) = \frac{1}{Z}\exp\left(-\frac{|s-\overrightarrow{\mu}_i|^2}{\overrightarrow{\sigma}_i^2} - \sum_{k'=1}^{m}\varphi_{k'}^{-1}(s)\lambda_a^{i,k'}\right) \tag{9}$$

with Z such as $\sum_{s\in\mathcal{S}} f_i(s) = 1$. Finally, the extrinsic LLR computation is done with the latest $\overrightarrow{\mu}$ and $\overrightarrow{\sigma}$ as

$$\lambda_e^{i,k} = \log\left(\frac{\sum_{s\in\mathcal{S}:\varphi_k^{-1}(s)=0} f_i(s)}{\sum_{s\in\mathcal{S}:\varphi_k^{-1}(s)=1} f_i(s)}\right) - \lambda_a^{i,k}. \tag{10}$$

The global receiver is a double loop receiver for which one or several successive updates (*auto-iterations*) of $m_{\text{equ}\to\text{dem}}$ and $m_{\text{dem}\to\text{equ}}$ can be applied within one global turbo iteration. The inner EP detector applying zero auto-iteration leads to a classical LMMSE detector, while the classic EP detector introduced by [7] is done with one auto-iteration with a reset of messages $m_{\text{equ}\to\text{dem}}$ between each global turbo iteration. Refined EP detection strategies have been proposed by [2], considering an extension to the case of more than one auto-iteration and no message reset between turbo iterations.

B. Belief Propagation based receivers

1) Standard Belief Propagation: A MIMO channel can be represented by a factor graph as given in Fig. 3 where variable nodes are associated with the N_t emitted symbols, and function nodes are referred to as the MIMO channel constraints at the N_r received antennas. Using this representation, the symbol BP algorithm can be applied, which consists in exchanging extrinsic messages (symbol probabilities or equivalent non-binary LLRs) between function and variable nodes using both a priori probabilities and channel likelihoods. At each iteration, for each pair of connected nodes, the algorithm updates the messages along the edges of the factor graph by computing two types of messages, referred to as α for messages exchanged between variable nodes and function nodes and β for messages exchanged between function and variable nodes. Each iteration starts with the computation of β messages followed by the computation of α messages.

In this paper, the messages are homogeneous to symbol LLRs, defined as the logarithm of the symbol probability minus the logarithm of the probability of a conventional reference symbol $\eta_{ref} \in \mathcal{S}$. First, the a priori symbol LLRs are given by $\forall i \in [\![1, N_t]\!], \forall \eta \in \mathcal{S},\ \lambda_a^i(\eta) = \log\left(\mathbb{P}\left(s_i = \eta\right)\right) - \log\left(\mathbb{P}\left(s_i = \eta_{ref}\right)\right)$. By definition, we have $\lambda_a^i(\eta_{ref}) = 0$. $\forall (i,j) \in [\![1, N_t]\!] \times [\![1, N_r]\!]$, and $\forall l \in [\![0, L]\!]$, we denote $\alpha_{i,j}^l$ (resp. $\beta_{i,j}^l$) the message vector of size $M-1$ exchanged at iteration l between the variable node i, corresponding to the i-th antenna and the function node j, corresponding to the MIMO channel constraint as observed at the j-th received antenna (resp. exchanged between the function node j and variable node i). For a given i, j and l we have, $\forall \eta \in \mathcal{S} \setminus \{\eta_{ref}\}$,

$$\alpha_{i,j}^l(\eta) = \lambda_a^i(\eta) + \sum_{t=1, t\neq j}^{N_r} \beta_{i,t}^l(\eta) \tag{11}$$

$$\beta_{i,j}^l(\eta) = \log\left(\frac{\sum_{\mathbf{z}\in\Psi_i^{\eta}} e^{-\frac{||\mathbf{y}_j - \mathbf{H}_j\mathbf{z}||^2}{N_0}} \prod_{i'=1, i'\neq i}^{N_t} e^{\alpha_{i',j}^{l-1}(z_{i'})}}{\sum_{\mathbf{z}\in\Psi_i^{\eta_{ref}}} e^{-\frac{||\mathbf{y}_j - \mathbf{H}_j\mathbf{z}||^2}{N_0}} \prod_{i'=1, i'\neq i}^{N_t} e^{\alpha_{i',j}^{l-1}(z_{i'})}}\right) \tag{12}$$

with $\mathbf{H}_j$ the j-th row of the matrix $\mathbf{H}$. Similarly to the MAP detector, the $log(sum(exp(\ .\ )))$ operation can be approximated by using a max to reduce the complexity. After L inner iterations, the output is computed as

$$\lambda_p^i(\eta) = \lambda_a^i(\eta) + \sum_{t=1}^{N_r} \beta_{i,t}^L(\eta). \tag{13}$$

In a MIMO system, there are $N_t N_r$ connections (a fully connected graph) which lead to a high number of iterations for convergence.

2) PIC Belief Propagation: Parallel Interference Cancellation (PIC) BP is a reduced complexity version of BP presented in [3]. It consists in making a PIC step at the beginning of each BP auto-iteration. Recall that the received signal on antenna j can be written as

$$\begin{aligned} y_j &= \sum_{i=1}^{N_t} h_{j,i}s_i + w_j \\ &= h_{j,i}s_i + \sum_{i'=1, i'\neq i}^{N_t} h_{j,i'}s_{i'} + w_j \end{aligned} \tag{14}$$

with $j \in [\![1, N_r]\!]$ the receiving antenna and $i \in [\![1, N_t]\!]$ the emitting antenna. A PIC is carried out by cancelling the sum term at the right-hand side of (14) using a soft symbol interference cancellation based on soft symbol estimates computed from α messages of the previous auto-iteration. The soft estimation of this sum term is computed for each edge between the emitting antenna i and the receiving antenna j and iteration l:

$$\hat{s}_{i,j}^l = \sum_{s'\in\mathcal{S}} s' \frac{e^{\alpha_{i,j}^l(s')}}{\sum_{s''\in\mathcal{S}} e^{\alpha_{i,j}^l(s'')}} \tag{15}$$

$$\sigma_{\hat{s}_{i,j}^l}^2 = \sum_{s'\in\mathcal{S}} ||s' - \hat{s}_{i,j}^l||^2 \frac{e^{\alpha_{i,j}^l(s')}}{\sum_{s''\in\mathcal{S}} e^{\alpha_{i,j}^l(s'')}}. \tag{16}$$

Once every soft symbol has been estimated, interference cancellation can be applied. Each β message update can be computationally complex because of the fully connected nature of the graph. To drastically decrease complexity, the number of variable nodes, connected to the updated function node, treated as in standard BP can be chosen very small compared to the complete set. An extreme choice is to keep

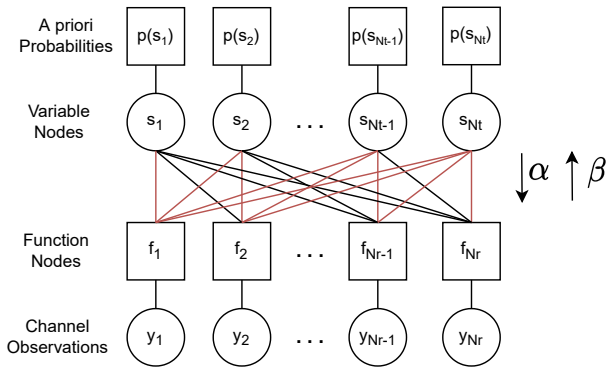

Fig. 3. Factor graph representation of the MIMO channel where red edges are remaining edges after QR pre-processing of the channel matrix.

only one edge, the edge of interest on which the message is sent as in [3], thus reducing the set of possible sent symbols from Ψ to $\mathcal{S}$. So when updating the β message between each connected function node j and variable node i, the updated channel observation on receiving antenna j and the updated noise variance associated with the node pair (i,j) (thermal noise + residual interference noise) becomes

$$\tilde{y}^l_{i,j} = y_j - \sum_{i'=1,i'\neq i}^{N_t} h_{j,i'}\hat{s}^l_{i',j} \tag{17}$$

$$\tilde{\sigma}^2_{\hat{w}^l_{i,j}} = N_0 + \sum_{i'=1,i'\neq i}^{N_t} \sigma^2_{\hat{s}^l_{i',j}}. \tag{18}$$

For a real constellation, the pseudo variance needs to be divided by 2. We apply the extreme option where only one variable node is selected, reducing as much as possible the computational complexity of β messages while keeping a minimal performance loss as shown in [3]. The β message update between every connected function node j and variable node i, $\forall \eta \in \mathcal{S} \setminus \{\eta_{ref}\}$, is done as

$$\beta^l_{i,j}(\eta) = -\frac{||\tilde{\mathbf{y}}^l_{i,j} - \mathbf{h}_i\eta||^2}{\sigma^2_{\hat{w}^l_{i,j}}} + \frac{||\tilde{\mathbf{y}}^l_{i,j} - \mathbf{h}_i\eta_{\text{ref}}||^2}{\sigma^2_{\hat{w}^l_{i,j}}}. \tag{19}$$

Note that the proposed PIC BP differs from [3] because α and β are updated at symbol level and not at bit level. The α update is carried out exactly as in classic BP as described in (11), i.e. as the sum of βs from every connected node, except the edge of interest, plus the a priori LLRs.

3) Proposed QR PIC Belief Propagation: To further improve the performance of the preceding algorithm, we propose QR PIC BP detection, which aims at combining the advantages of QR [8] and PIC processing. QR PIC BP requires computing less a posteriori soft symbol estimates and fewer multiplications for β messages than classic BP or PIC BP approaches. The first step consists in pre-conditioning the factor graph thanks to a QR decomposition of $\mathbf{H}$ as shown in (20). The second step is to generate a posteriori soft symbols for every variable node still connected in the updated graph (remaining red edges in Fig. 3) and the update of the remaining βs and αs is done.

TABLE I
TABLE OF OPERATIONS COST [9]

Operations	Symbols	Mult.	Add.	FLOPs
Real Addition/Subtraction	A_r	0	1	0.5
Real Multiplication/Division	M_r	1	0	0.5
Comparison	C_r	0	0	0.5
Complex-Real Mult./Div.	M_{rc}	2	0	1
Complex Addition	A_c	0	2	1
Complex Mult.	M_c	4	2	3
Complex Div.	D_c	8	3	5.5
Real Square Root	S_r	0	0	3
Squared Complex Norm	P_c	2	1	1.5
Exponential [10]	X_e	1	4	2.5
Memory Access	M_a	0	0	0

The QR pre-processing on the channel matrix is done to reduce the number of non-zero elements. $\mathbf{H} = \mathbf{QR}$ decomposes the channel matrix into an upper triangular matrix $\mathbf{R}$ and a unitary matrix $\mathbf{Q}$. The observation model can be updated as

$$\mathbf{y} \leftarrow \mathbf{Q}^H\mathbf{y} \qquad \text{and} \qquad \mathbf{H} \leftarrow \mathbf{R} \tag{20}$$

and PIC BP is applied to the new observation and channel matrix. Statistical properties of the additive noise remain unchanged due to the unitary property of $\mathbf{Q}$. QR PIC BP works on a less connected graph, with only $N_t(N_t+1)/2$ edges instead of N_tN_r edges of the original graph. There are fewer messages exchanged which results in a even less complex algorithm than PIC BP.

IV. Complexity Analysis

A. Computational Complexity

We now analyse the complexities of the preceding algorithms, which scale very differently. On one hand, the MAP complexity is dominated by the need of comparing the probability of each multi-dimensional symbol of Ψ with the channel observations. The algorithm complexity is of the order of $\mathcal{O}(M^{N_t}N_tN_r)$.

Standard symbol BP also relies on joint MIMO symbol detection (see (12)), which is the complexity main contributor. In addition, for each of the L auto-iterations, there are N_tN_rM messages exchanged (i.e. α and β updates). So, standard BP has a complexity of the order of $\mathcal{O}(M^{N_t}N_tN_r + N_tN_rML)$.

PIC BP has a much lower overall complexity than MAP or standard BP since it does not require to work with the set Ψ but only with the set $\mathcal{S}$ when it comes to the β messages computation. The number of required auto-iterations is lower because of the reduced number of selected edges. The complexity of the BP part is impressively reduced to $\mathcal{O}(N_tN_rML)$. The cost of the additional PIC process is of the same order as the BP part.

QR PIC BP requires less soft symbol generation and less BP message exchanges thanks to the QR pre-processing. The overall cost of QR PIC BP is $\mathcal{O}(N_t(N_t+1)ML)$. The cost of the QR factorisation is much smaller than the cost of

TABLE II
NUMBER OF AUTO-ITERATIONS FOR EACH RECEIVER.
THE FORMAT IS: $\{2 \times 2$ MIMO$\}$ - $\{4 \times 4$ MIMO$\}$ AUTO-ITERATIONS.

Modulation	EP	BP	QR BP	PIC BP	QR PIC BP
BPSK	1 - 1	2 - 6	1 - 3	1 - 4	1 - 1
QPSK	1 - 1	2 - 6	1 - 3	1 - 4	1 - 2
8PSK	1 - 1	2 - 6	1 - 3	2 - 4	1 - 2
16QAM	1 -	2 -	1 -	2 -	1 -
32QAM	1 -	2 -	1 -	4 -	1 -
64QAM	1 -	2 -	1 -	4 -	1 -

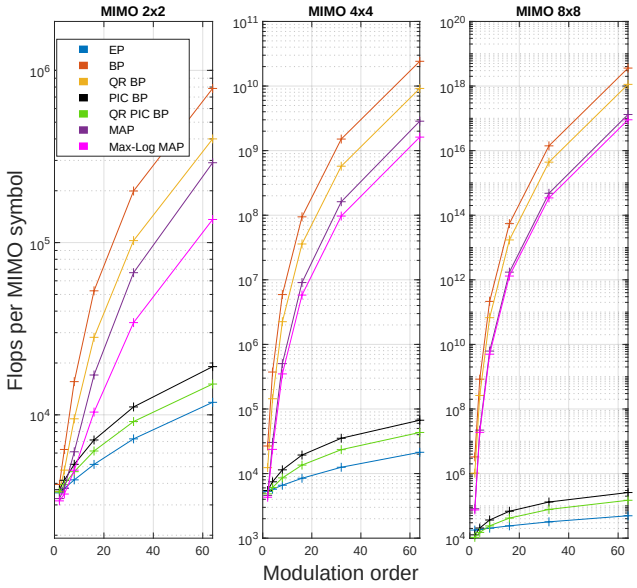

Fig. 4. Complexity function of the modulation order for 2×2 , 4×4 and 8×8 MIMO with a convolutional code $[(133, 171)_8, R = 1/2]$ without turbo iteration.

PIC BP. Finally, EP's complexity is dominated by the matrix inversion done in (5) with a complexity of the order of $\mathcal{O}(N_t^3)$. The LLR computation (10) has a complexity of the order of $\mathcal{O}(MN_t)$, which is way lower than the one required for a matrix inversion.

Besides the previous asymptotic complexity behaviours, we have conducted a deeper computational complexity analysis. First, we have written an efficient software implementation of each given algorithm. Then, the number of occurrences of each elementary operation in the implementation has been counted. Finally, a reasonable weight in FLOPs has been assigned to each elementary operation (see Table. I). The cost of the exponential operation X_e is estimated at $X_e = M_r + 4A_r$ based on the analysis of the implementation proposed in [10]. In this way, it is possible to associate a unique computational cost per algorithm, as a function of its parameters, which has been used to produce all the complexity figures in this section. Fig. 4 shows the complexity of every presented algorithm for several MIMO contexts and constellation size using the unique computational cost per algorithm.

B. Complexity/Performance Trade-off

In this section, we compare the different complexity/performance points achieved by BP-based and EP detectors for a target coded BER. Random interleaving has been used for every simulation. The number of auto-iterations of each algorithm, reported in Tab. II, has been determined by numerical trials (not reported here) in order to let the algorithm converge to its best performance. In the complexity vs SNR figures, each curve represents the trade-off for a given algorithm applied to a specific modulation, coding and MIMO scheme. Every point of a given curve shows the trade-off at a different turbo iteration of the associated detector. The lowest point (in terms of complexity) of each curve corresponds to turbo iteration 0.

Simulations shown in Fig. 5 are done using a convolutional code with polynomials $(133, 171)_8$ and rate $1/2$. After random interleaving, a codeword is mapped to 1000 MIMO symbols, which are then sent through an ergodic spatially uncorrelated Rayleigh channel, known at the receiver side. The maximum number of turbo iterations is set to 8 in this 2×2 MIMO context. For small constellations (BPSK and QPSK), MAP is the least complex algorithm with the best performance. When the constellation order starts growing (e.g. 8PSK, 16QAM, 32QAM, 64QAM), MAP, BP and QR BP become too complex compared to EP, PIC BP and the proposed QR PIC BP. We can note that PIC BP has a worst performance than any other detector. Finally, EP and QR PIC BP have almost the same complexity/performance trade-off.

Fig. 6 and Fig 7 shows results on the same 2×2 MIMO system, but with a LDPC code of length $K = 1944$ and rate $R = 1/2$ and $R = 3/4$ respectively. Turbo iterations with LDPC have a much lower impact on performance than with convolutional code, due to the different extrinsic transfer functions of the two codes [9], even with an higher code rate. However, Fig. 6 shows that the conclusions found with a convolutional code hold true also for the LDPC case. Fig. 7 shows on one hand that, with small spectral efficiency modulations, the MAP and QR BP achieve better performance than EP, QR PIC BP and PIC BP. On the other hand, with high spectral efficiency modulations, the EP and QR PIC BP achieve the same performance of MAP and QR BP while requiring much smaller complexity.

The complexity/performance trade-off of a 4×4 MIMO system using a convolutional code $(133, 171)_8$ with codewords mapped to 1000 MIMO symbols is shown in Fig. 8. For the smallest constellation (i.e. BPSK) MAP is still the best algorithm because of its optimal performance and lowest complexity. Then, for QPSK and 8PSK, EP and QR PIC BP have the lowest complexity and similar performance (MAP, BP and QR BP were too complex to be simulated with higher constellations). Finally, Fig. 9 shows the behaviour of the algorithms in a large 16×16 MIMO system. Without turbo iterations, EP is the best performing algorithm. With turbo iterations, EP and QR PIC BP converge to the same performance. Hence, QR PIC BP seems a promising candidate also for application in massive MIMO systems.

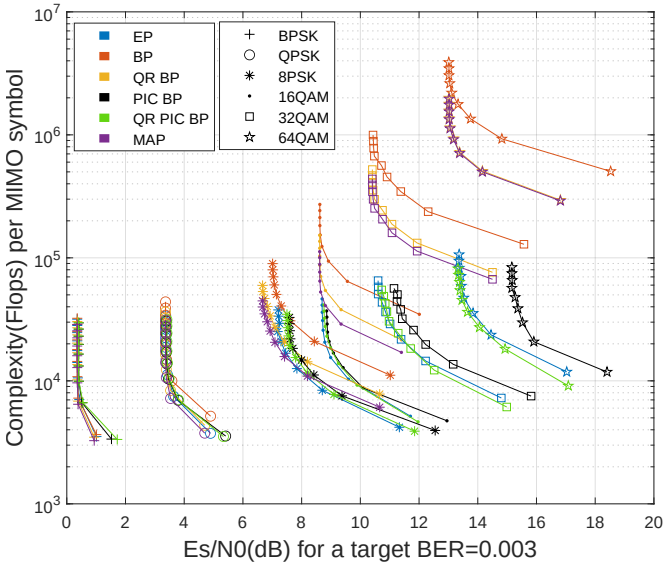

Fig. 5. Complexity for a target BER of 10^{-3} in a 2×2 MIMO system with a convolutional code $(133, 171)_8$ and rate $1/2$.

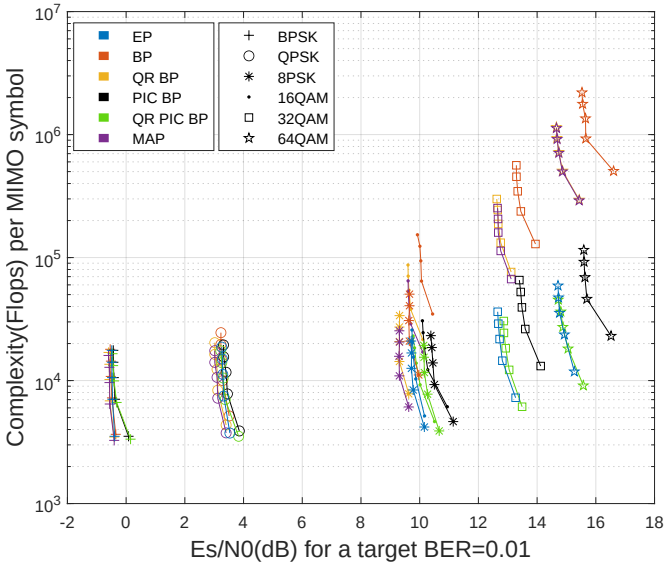

Fig. 6. Complexity for a target BER of 10^{-2} in a 2×2 MIMO system using a LDPC of length 1944 and rate $1/2$.

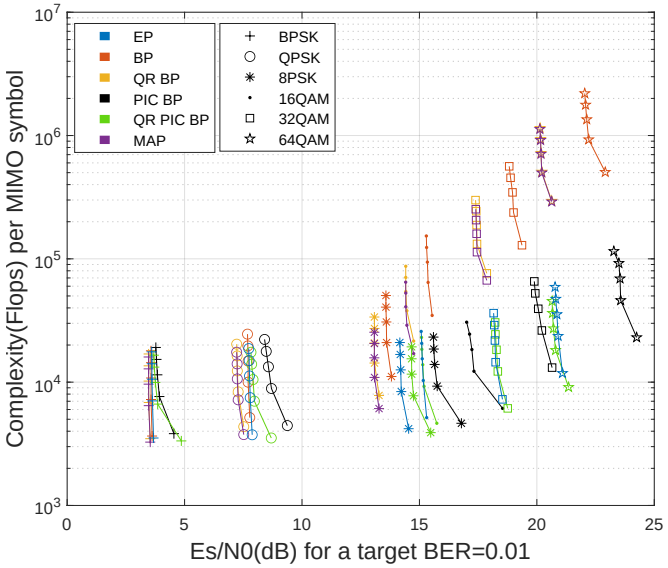

Fig. 7. Complexity for a target BER of 10^{-2} in a 2×2 MIMO system using a LDPC of length 1944 and rate $3/4$.

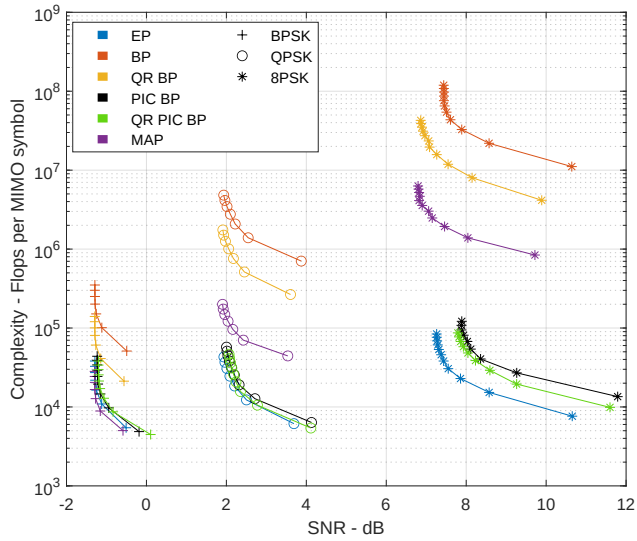

Fig. 8. Complexity for a target BER of $3 * 10^{-2}$ in a 4×4 MIMO system with a convolutional code $(133, 171)_8$ and rate $1/2$.

C. Symmetric Information Rate

In this section, we compare the symmetric information rate (SIR) of the low complexity detectors to the coded modulation capacity. We introduce the coded modulation (CM) as the best achievable constrained capacity. It is calculated as the mutual information between the estimated MIMO symbol with the MAP criterion and the sent symbol $x \in \Psi$. We define the BICM capacity as the mutual information between estimated bits at the output of the detector, after marginalisation, and the actual sent bits. Fig.10 shows the SIR using BICM of the different detectors for a2×2 MIMO system with a 64QAM modulation. Each detector uses the same number of auto-iterations as described in Table II. The best achievable rate is 12 bits per channel use (2 antennas and 6 bits per symbol). To achieve a SIR of 6 bits per channel use, one should use a forward Error Correcting (FEC) code of rate 1/2, where EP, QR BP and QR PIC BP can deliver 6 bits per channel use at 13.5 dB. MAP algorithm (name Coded Modulation - CM here) can do it at 11.5 dB while BP and PIC BP can only do it at 14.3 dB. We notice that PIC BP cannot achieve the upper bound of capacity (it needs 20 auto-iterations to do so) whereas QR PIC BP achieves it with only 1 auto-iteration. In this MIMO V2V-NR context, EP and QR PIC BP have the same SIR, no matter the code rate, which makes them good competitors for this kind of vehicular applications,

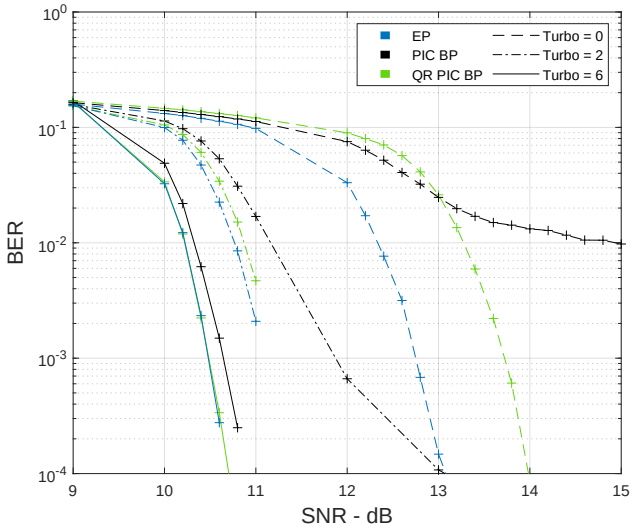

Fig. 9. Performances of turbo iterated EP, PIC BP and QR PIC BP in a 16×16 MIMO with 16QAM using a LDPC of length $K = 4096$ and rate $R = 1/2$ with auto-iterations: [EP = 2, PIC BP = 10, QR PIC BP = 4].

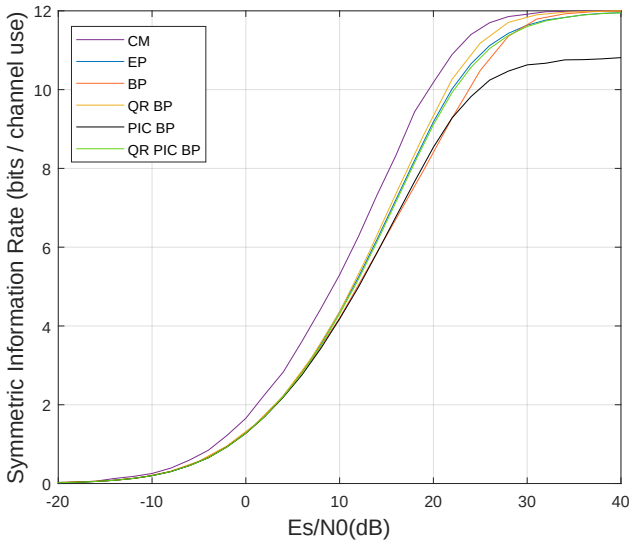

Fig. 10. SIR of low complexity receivers and coded modulation in MIMO 2×2 64QAM.

as seen in section IV-B for their similar complexity. Those SIR performances are achievable only if we use a optimally designed FEC code for each detector. Turbo-iterations are the way of closing the gap between CM and the other detectors.

V. Conclusion

This study has identified two clear zones (or sets) of MIMO and system configurations, where the selection of the most appropriate algorithm on a complexity/performance trade-off basis is different. The first highlighted zone is for small MIMO systems (2×2 MIMO with BPSK, QPSK and 8PSK - 4×4 MIMO with BPSK). In these scenarios, the algorithm with the best complexity/performance trade-off is (Max-Log) MAP. EP, PIC BP and QR PIC BP have a slightly less interesting complexity/performance trade-off than MAP, whereas BP and QR BP are much more complex. The second zone is for more complex MIMO systems (2×2 MIMO with 16QAM, 32QAM and 64QAM - 4×4 MIMO with QPSK and 8PSK). In this zone, MAP becomes obviously much more complex than in the first area, while EP, PIC BP and QR PIC BP do not have much higher complexity. In most cases, EP and QR PIC BP have a substantially equivalent complexity/performance trade-off, whereas PIC BP is slightly less performing. One should consider using EP or the proposed QR PIC BP when it comes to those MIMO systems. In the V2V-NR [5] scenario, MIMO 2×2 64QAM, EP and QR PIC BP are much less complex and achieve great performance which results in an advantageous performance/complexity trade-off.

Moreover, we have shown that the proposed QR PIC BP algorithm performs similarly to EP in MIMO contexts up to 16×16. Investigating the behaviour of QR PIC BP in mMIMO systems seems a very interesting perspective, due to its low computational cost. Another path of future research is to compare QR PIC BP with simplified EP algorithms and/or low-complexity AMP algorithms.

References

[1] R. W. Heath and A. Lozano. *Foundations of MIMO communication*. Cambridge University Press, 2019.

[2] J. J. Murillo-Fuentes et al. "A Low-Complexity Double EP-Based Detector for Iterative Detection and Decoding in MIMO". In: *IEEE Trans. Commun.* 69.3 (Mar. 2021).

[3] W. Fukuda et al. "Low-Complexity Detection Based on Belief Propagation in a Massive MIMO System". In: *IEEE 77th VTC Spring*. June 2013, pp. 1–5.

[4] S. Wu et al. "Low-Complexity Iterative Detection for Large-Scale Multiuser MIMO-OFDM Systems Using Approximate Message Passing". In: *IEEE J. Sel. Topics Signal Process* 8.5 (2014), pp. 902–915.

[5] Shao-Yu Lien et al. "3GPP NR Sidelink Transmissions Toward 5G V2X". In: *IEEE Access* 8 (2020).

[6] Steven M Kay. *Fundamentals of statistical signal processing: estimation theory*. Prentice-Hall, Inc., 1993.

[7] M. Senst et al. "How the Framework of Expectation Propagation Yields an Iterative IC-LMMSE MIMO Receiver". In: IEEE Globecom. Dec. 2011, pp. 1–6.

[8] S. Park and S. Choi. "QR decomposition aided belief propagation detector for MIMO systems". In: *Electronics Letters* 51.11 (May 2015), pp. 873–874.

[9] S. Sahin. "Advanced receivers for distributed cooperation in mobile ad hoc networks". INP Toulouse, 2019.

[10] N. N. Schraudolph. "A Fast, Compact Approximation of the Exponential Function". In: *Neural Computation* 11.4 (May 1999), pp. 853–862.