This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE Globecom 2010 proceedings.

Variational Message-Passing for Joint Channel Estimation and Decoding in MIMO-OFDM

Gunvor Elisabeth Kirkelund*, Carles Navarro Manchón*, Lars P. B. Christensen†, Erwin Riegler§ and Bernard Henri Fleury*
*Department of Electronic Systems, Aalborg University, Denmark
†Modem Algorithm Design, NOKIA, Denmark
§Vienna University of Technology (VUT)
Email: {gunvor, cnm}@es.aau.dk, lars.christensen@nokia.com, erwin.riegler@nt.tuwien.ac.at, bfl@es.aau.dk

***Abstract*—In this contribution, a multi-user receiver for M-QAM MIMO-OFDM operating in time-varying and frequency-selective channels is derived. The proposed architecture jointly performs semi-blind estimation of the channel weights and noise inverse variance, serial interference cancellation and decoding in an iterative manner. The scheme relies on a variational message-passing approach, which enables a joint design of all these functionalities or blocks but the last one. Decoding is performed using the sum-product algorithm. This is in contrast to nowadays proposed approaches in which all these blocks are designed and optimized individually. Simulation results show that the proposed receiver outperforms in coded bit-error-rate a state-of-the-art iterative receiver of same complexity, in which all blocks are designed independently. Joint block design and, as a result, the fact that the uncertainty in the channel estimation is accounted for in the proposed receiver explain this better performance.**

I. INTRODUCTION

During recent years, algorithms based on iterative information processing or "turbo" techniques have become widespread in wireless receiver design [1]–[3]. The success of these algorithms can be explained by their remarkable properties: high performance at tractable complexity and flexibility in their design. An emblematic example is turbo-codes, which, when associated with turbo-decoding, allow for transmission close to capacity at tractable complexity [1].

In this paper, we focus on a specific application of iterative information processing, namely to design efficient, feasible algorithms for channel estimation (i.e. estimation of both the channel transfer function and the channel inverse noise variance), interference cancellation, and decoding in MIMO-OFDM systems. Some related work is already available in the literature. Worth noticing is the iterative algorithm for detection and interference cancellation [4] applied to multiuser CDMA. This algorithm is extended for various transmission schemes in [5]–[7] to include estimation of the channel response into the iterative process. We coin this receiver the LMMSE-based receiver, according to the dominant structure implemented in its constituent blocks. An essential feature of this receiver is that its constituent blocks are designed and optimized individually. These blocks are connected afterwards to form the iterative structure.

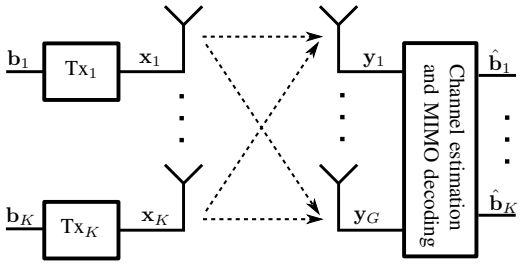

Fig. 1. Baseband signal model of the considered MIMO-OFDM system.

In this contribution, we apply variational Bayesian (VB) inference [8] and one of its applications, namely the variational Bayesian expectation maximization (VBEM) algorithm [9] to perform channel weight and noise inverse variance estimation as well as serial interference cancellation in an M-QAM MIMO-OFDM system operating in time-variant frequency-selective channels. Decoding is performed using the sum-product (SP) algorithm [3]. The VBEM algorithm has already been applied in [10] for GSM channel estimation and detection. In [11] it is combined with the sum-product algorithm for the design of a multiuser CDMA receiver. Further related work is found in [12]–[15]. In our paper, we apply the VBEM scheme in [11] to MIMO-OFDM and reformulate it as a variational message-passing (VMP) algorithm on factor graphs [16].

The proposed VMP receiver and the LMMSE-based receiver from [5]–[7] share similar features in their respective structures. Thus, we find it useful to also include a comparison of the two schemes. A crucial difference is that the estimation of the noise and residual interference power in the VMP receiver accounts for the uncertainty in the channel coefficient estimates, an effect not considered in the LMMSE-based receiver. This, combined with the joint design of all receiver blocks but decoding, yields a superior performance of the VMP receiver, as our simulation results demonstrate.

The notational convention for the rest of the paper is as follows: the superscripts $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ denote transposition and Hermitian transposition respectively. The symbol $\cdot \propto \cdot$ denotes proportionality. The trace operator is designated as $\mathrm{tr}(\cdot)$. The expectation operation with respect to a function $q(x)$ is represented by $\langle\cdot\rangle_{q(x)}$. The newest estimate of the mean or covariance of a variable is denoted by $\hat{\cdot}$. The operators $\mathrm{diag}(\cdot)$ and $\mathrm{Diag}(\cdot)$ denote the vectorized diagonal of a matrix and the diagonalized matrix of a vector respectively. For matrices $\mathbf{A}$ and $\mathbf{B}$, the Kronecker and Hadamard products are represented

978-1-4244-5638-3/10/$26.00 ©2010 IEEE

by $\mathbf{A}\otimes\mathbf{B}$ and $\mathbf{A}\odot\mathbf{B}$ respectively (for the Hadamard product $\mathbf{A}$ and $\mathbf{B}$ are assumed to have the same dimension). The identity matrix of dimension K is designated as $\mathbf{I}_K$ and $\mathbf{1}_G$ represents the all-one matrix of dimension $G \times G$. We employ $\mathbf{0}_K$ and $[1 \dots 1]_K$ to designate respectively the all-zero column-vector and the all-one row-vector of length K.

II. Signal Model

We consider the LTE-like MIMO-OFDM system depicted in Fig. 1 in which we have K transmitters, indexed by k, and G receivers, indexed by g. In the kth transmitter, denoted by Tx_k, the bit-stream $\mathbf{b}_k$ is encoded, interleaved and modulated into data symbols, which are then multiplexed with pilot symbols to allow for channel estimation in the receiver. Pilot and data symbols are arranged in an OFDM frame of L OFDM symbols consisting of N subcarriers each. The OFDM frame of Tx_k is represented by $\mathbf{x}_k \triangleq [x_{k11} \dots x_{knl} \dots x_{kNL}]^{\mathrm{T}} \in \mathcal{X}_k$, where l indexes the OFDM symbols and n indexes the subcarrier number. The set $\mathcal{X}_k$ of legal M-ary sequences of Tx_k is determined by the coding and modulation scheme and the multiplexing scheme of data and pilot symbols.

The OFDM frames are transmitted across a time-variant frequency-selective channel. The samples of the time-frequency response of the sub-channel from transmit antenna k to receive antenna g are concatenated in the channel weight vector $\mathbf{a}_{gk} \triangleq [a_{gk11} \dots a_{gk1L} \dots a_{gknl} \dots a_{gkNL}]^{\mathrm{T}}$. Assuming that inter-symbol and inter-subcarrier interferences are negligible, the received signals at all G antenna ports are given in vector notation by

$$\mathbf{y} = \sum_{k=1}^{K} \mathbf{A}_k \mathbf{x}_k + \mathbf{w} \tag{1}$$

$$= \mathbf{X}\mathbf{a} + \mathbf{w} \tag{2}$$

$$= \mathbf{A}\mathbf{x} + \mathbf{w}. \tag{3}$$

The vector $\mathbf{y}$ is the concatenation of the output vectors of all receive antennas, $\mathbf{y} \triangleq [\mathbf{y}_1^{\mathrm{T}} \dots \mathbf{y}_g^{\mathrm{T}} \dots \mathbf{y}_G^{\mathrm{T}}]^{\mathrm{T}}$ with $\mathbf{y}_g \triangleq [y_{g11} \dots y_{gnl} \dots y_{gNL}]^{\mathrm{T}}$ denoting the output of receive antenna g. The channel matrix for transmitter k is defined as $\mathbf{A}_k \triangleq \mathrm{Diag}(\mathbf{a}_k)([1 \dots 1]_G^{\mathrm{T}} \otimes \mathbf{I}_N)$. The noise vector $\mathbf{w}$ is white and circularly symmetric complex Gaussian: $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}_{GNL}, \sigma_w^2 \mathbf{I}_{GNL})$, with σ_w^2 denoting the noise variance. We define the precision parameter $\lambda \triangleq \sigma_w^{-2}$. The matrix $\mathbf{X}$ is defined as $\mathbf{X} \triangleq \mathbf{I}_G \otimes (([1 \dots 1]_K \otimes \mathbf{I}_N)\mathrm{Diag}(\mathbf{x}))$ and $\mathbf{a} \triangleq [\mathbf{a}_{11}^{\mathrm{T}} \dots \mathbf{a}_{gk}^{\mathrm{T}} \dots \mathbf{a}_{GK}^{\mathrm{T}}]^{\mathrm{T}}$. The matrix $\mathbf{A} \triangleq (\mathbf{I}_G \otimes ([1 \dots 1]_K \otimes \mathbf{I}_N))\mathrm{Diag}(\mathbf{a})(([1 \dots 1]_G^{\mathrm{T}} \otimes \mathbf{I}_K) \otimes \mathbf{I}_N)$ is the MIMO channel matrix. The vector $\mathbf{x} \triangleq [\mathbf{x}_1^{\mathrm{T}} \dots \mathbf{x}_k^{\mathrm{T}} \dots \mathbf{x}_K^{\mathrm{T}}]^{\mathrm{T}}$ contains the concatenated OFDM codewords from all transmit antennas. The receiver outputs an estimate $\hat{\mathbf{b}}_k$ of the bit-stream for any k.

III. Graphical Representation

In this section, we present a graphical representation of the signal model introduced in the previous section. This graphical representation will be used to derive the message-passing algorithm in Section IV. Let

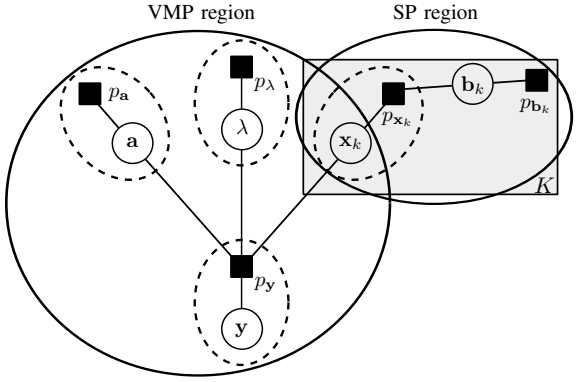

Fig. 2. Factor-graph [3] of the signal model in Section II. The parameter K indicates that the corresponding block is repeated K times, one for each transmitter. Notice that the left region is equivalent to the Bayesian network representation from [17].

$$\mathbf{\Phi} \triangleq \{\mathbf{y}, \mathbf{a}, \lambda, \mathbf{x}_1, \dots \mathbf{x}_K, \mathbf{b}_1, \dots \mathbf{b}_K\} \tag{4}$$

denote the set of all (observed and unobserved) variables in (1). Based on the assumptions made in Section II, the joint probability density function (pdf) of $\mathbf{\Phi}$ factorizes as

$$p_{\mathbf{\Phi}}(\mathbf{\Phi}) = p_{\mathbf{y}}(\mathbf{y}|\mathbf{a}, \lambda, \mathbf{x}_1, \dots \mathbf{x}_K) p_{\mathbf{a}}(\mathbf{a}) p_\lambda(\lambda) \prod_k p_{\mathbf{x}_k}(\mathbf{x}_k|\mathbf{b}_k) p_{\mathbf{b}_k}(\mathbf{b}_k). \tag{5}$$

The constraints imposed by coding, modulation and multiplexing of the deterministic pilot symbols are included in the factor $p_{\mathbf{x}_k}$ for transmitter k. A straightforward graphical representation of this factorization is the Tanner factor-graph [3] depicted in Fig. 2. Factors are represented as squares, variables as circles. An edge between a variable node and a factor node indicates that the variable is an argument of the factor.

Based on this graphical representation of the signal model, we employ iterative algorithms to estimate the joint pdf $p_{\mathbf{\Phi}}$. We split the graph into two regions as depicted in Fig. 2. In the right-hand region, we apply the SP algorithm [3] to compute the marginals $p_{\mathbf{x}_k}$ and $p_{\mathbf{b}_k}$. In the left-hand region, we apply the VMP algorithm [16] to estimate $p_{\mathbf{a}}$ and p_λ. The VMP algorithm is used to reformulate the VB inference method proposed in [11] in terms of messages.

The motivation for splitting the Tanner graph in this way and applying two different message-passing methods is as follows. The SP algorithm is a well-established algorithm for computing the marginal probability mass functions $p_{\mathbf{x}_k}$ and $p_{\mathbf{b}_k}$ in known channel conditions. Direct computation of the channel marginals $p_{\mathbf{a}}$ and p_λ by means of the SP algorithm is, however, computationally infeasible. In this case, one has to rely on techniques for approximating these marginals, e.g. particle filters or the EM algorithm [18]. Here, we propose another avenue and compute these marginals with the VMP algorithm. We define the set of unknown variables in the VMP region as $\mathbf{\Phi}_{\mathrm{VMP}} \triangleq \{\mathbf{a}, \lambda, \mathbf{x}_1, \dots \mathbf{x}_K\} \subseteq \mathbf{\Phi}$.

978-1-4244-5638-3/10/$26.00 ©2010 IEEE

Variational Message-Passing (VMP)

We consider an arbitrary factor-graph. The message from factor node f to a variable node ϕ in the set $\mathcal{N}_f$ of neighbouring nodes of f is

$$m_{f\to\phi} \triangleq \exp\langle \ln f\rangle_{m_{\phi'\to f}\forall\phi'\in\mathcal{N}_f\setminus\phi}. \quad (6)$$

The message from variable node ϕ to any factor node f in the set $\mathcal{N}_\phi$ of factor nodes neighbouring ϕ is

$$m_{\phi\to f} \triangleq \prod_{f'\in\mathcal{N}_\phi} m_{f'\to\phi}. \quad (7)$$

The estimated auxiliary function of ϕ is

$$b_\phi \propto m_{\phi\to f}. \quad (8)$$

IV. Variational Message-Passing

In this section, we apply the VMP algorithm [11], [17] to the left-hand region in the factor-graph in Fig. 2, see [16] and references therein for variational inference on factor-graphs. The message-passing rules are summarized in (6)-(8). Their derivations are sketched in App. A.

The VMP algorithm approximates the joint pdf $p_{\mathbf{\Phi}_{\mathrm{VMP}},\mathbf{y}}(\mathbf{\Phi}_{\mathrm{VMP}},\mathbf{y}) = p_{\mathbf{y}}(\mathbf{y}|\mathbf{a},\lambda,\mathbf{x}_1,\dots\mathbf{x}_K)p_{\mathbf{a}}(\mathbf{a})p_\lambda(\lambda)\prod_k p_{\mathbf{x}_k}(\mathbf{x}_k)$ with an auxiliary function $b_{\mathbf{\Phi}_{\mathrm{VMP}}}(\mathbf{\Phi}_{\mathrm{VMP}})$ in such a way that the KL divergence from $b_{\mathbf{\Phi}_{\mathrm{VMP}}}(\mathbf{\Phi}_{\mathrm{VMP}})$ to $p_{\mathbf{\Phi}_{\mathrm{VMP}},\mathbf{y}}(\mathbf{\Phi}_{\mathrm{VMP}},\mathbf{y})$ is minimized [17]. We constrain the auxiliary function to factorize according to $b_{\mathbf{\Phi}_{\mathrm{VMP}}}(\mathbf{\Phi}_{\mathrm{VMP}}) = b_{\mathbf{a}}(\mathbf{a})b_\lambda(\lambda)b_{\mathbf{x}_1}(\mathbf{x}_1)\dots b_{\mathbf{x}_K}(\mathbf{x}_K)$.

The VMP algorithm implements sequential message updates to update the factors in $b_{\mathbf{\Phi}_{\mathrm{VMP}}}(\mathbf{\Phi}_{\mathrm{VMP}})$. Updating $b_{\mathbf{a}}(\mathbf{a})$, $b_\lambda(\lambda)$, and $b_{\mathbf{x}_k}(\mathbf{x}_k)$ corresponds to estimating the channel weights, estimating the precision parameter, i.e. the channel inverse noise variance, and interference cancellation, respectively.

A. Estimation of the Channel Weights

In this subsection, we derive the messages to and from the variable node $\mathbf{a}$. These messages are used to update $b_{\mathbf{a}}$ by means of (8). The message to node $\mathbf{a}$ from $p_{\mathbf{y}}$ is obtained from (6):

$$m_{p_{\mathbf{y}}\to\mathbf{a}} = \exp\left(\langle \ln\, p_{\mathbf{y}}(\mathbf{y}|\mathbf{a},\lambda,\mathbf{x})\rangle_{m_{\lambda\to p_{\mathbf{y}}}\prod_k m_{\mathbf{x}_k\to p_{\mathbf{y}}}}\right). \quad (16)$$

Solving the expectation yields

$$m_{p_{\mathbf{y}}\to\mathbf{a}} \propto p_{\mathcal{CN}}(\hat{\lambda}_{\mathrm{VMP}}\hat{\mathbf{C}}_{p_{\mathbf{y}}\to\mathbf{a}}\hat{\mathbf{X}}^{\mathrm{H}}\mathbf{y}, \hat{\mathbf{C}}_{p_{\mathbf{y}}\to\mathbf{a}}). \quad (17)$$

Here, $p_{\mathcal{CN}}(\boldsymbol{\mu},\mathbf{C})$ is a multivariate complex Gaussian pdf with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$, and $\hat{\mathbf{C}}_{p_{\mathbf{y}}\to\mathbf{a}} \triangleq (\hat{\lambda}_{\mathrm{VMP}}\hat{\mathbf{X}}^{\mathrm{H}}\hat{\mathbf{X}} + \hat{\lambda}_{\mathrm{VMP}}(\mathbf{I}_G\otimes\hat{\mathbf{C}}_{\mathbf{x}}))^{-1}$. The matrix $\hat{\mathbf{C}}_{\mathbf{x}}$ is the block-diagonal concatenation of the estimates $\hat{\mathbf{C}}_{\mathbf{x}_k}$ of the covariance matrices of $\mathbf{x}_k$, $k = 1\dots K$. Both $\hat{\mathbf{C}}_{\mathbf{x}_k}$ and the estimate $\hat{\lambda}_{\mathrm{VMP}}$ of the precision parameter are defined later in this section. We impose the prior $p_{\mathbf{a}}$ to belong to the family of conjugate pdfs of $\mathbf{a}$ for $p_{\mathbf{y}}$. This choice guarantees that the auxiliary pdf $b_{\mathbf{a}}$ is also in this family. From (16) the conjugate family of pdfs of $\mathbf{a}$ for $p_{\mathbf{y}}$ is the Gaussian family. Thus, from (6)

$$m_{p_{\mathbf{a}}\to\mathbf{a}} = p_{\mathcal{CN}}(\mathbf{0}_{GKNL}, \mathbf{C}_{\mathbf{a}}), \quad (18)$$

where $\mathbf{C}_{\mathbf{a}}$ is the prior channel covariance matrix. Inserting (17) and (18) in (7) yields

$$m_{\mathbf{a}\to p_{\mathbf{y}}} \propto p_{\mathcal{CN}}(\hat{\mathbf{a}}, \hat{\mathbf{C}}_{\mathbf{a}}) = b_{\mathbf{a}} \quad (19)$$

with $\hat{\mathbf{a}} = \hat{\lambda}_{\mathrm{VMP}}\hat{\mathbf{C}}_{\mathbf{a}}\hat{\mathbf{X}}^{\mathrm{H}}\mathbf{y}$ and $\hat{\mathbf{C}}_{\mathbf{a}} = (\mathbf{C}_{\mathbf{a}}^{-1} + \hat{\mathbf{C}}_{p_{\mathbf{y}}\to\mathbf{a}}^{-1})^{-1}$. As the Gaussian pdf is fully defined by these two moments – its natural statistics – it is enough to pass them to $p_{\mathbf{y}}$.

B. Estimation of the Precision Parameter

In this subsection, we define the messages to and from variable node λ. The auxiliary function b_λ is then updated by plugging these messages in (8). The message from $p_{\mathbf{y}}$ to λ reads from (6)

$$m_{p_{\mathbf{y}}\to\lambda} = \exp\left(\langle \ln\, p_{\mathbf{y}}(\mathbf{y}|\mathbf{a},\lambda,\mathbf{x})\rangle_{m_{\mathbf{a}\to p_{\mathbf{y}}}\prod_k m_{\mathbf{x}_k\to p_{\mathbf{y}}}}\right). \quad (20)$$

Evaluating the expectation under the assumption that the messages $m_{\mathbf{a}\to p_{\mathbf{y}}}$ and $m_{\mathbf{x}_k\to p_{\mathbf{y}}}$, $k = 1\dots K$, are Gaussian densities [11] yields

$$m_{p_{\mathbf{y}}\to\lambda} \propto p_{\mathcal{CW}_1}\left(\hat{W}^{-1}, GNL+1\right). \quad (21)$$

In this expression, $p_{\mathcal{CW}_F}(\mathbf{M}^{-1},d)$ is a complex Wishart pdf defined by three parameters: the dimension F, the degree of freedom d, and a matrix $\mathbf{M}$ of dimension $F\times F$ [19]. Here, $F=1$, $d = GNL+1$, and $\mathbf{M}$ is a scalar given as $\hat{W} \triangleq \mathrm{tr}((\mathbf{y}-\hat{\mathbf{A}}\hat{\mathbf{x}})(\mathbf{y}-\hat{\mathbf{A}}\hat{\mathbf{x}})^{\mathrm{H}} + \hat{\mathbf{X}}\hat{\mathbf{C}}_{\mathbf{a}}\hat{\mathbf{X}}^{\mathrm{H}} + \sum_k \hat{\mathbf{A}}_k\hat{\mathbf{C}}_{\mathbf{x}_k}\hat{\mathbf{A}}_k^{\mathrm{H}} + \sum_k(\mathbf{1}_G\otimes\hat{\mathbf{C}}_{\mathbf{x}_k})\odot\mathrm{Diag}(\mathrm{diag}(\hat{\mathbf{C}}_{\mathbf{a}_k})))$. The estimate $\hat{\mathbf{C}}_{\mathbf{a}_k}$ of the auto-covariance matrix of $\mathbf{a}_k$ can be obtained from $\hat{\mathbf{C}}_{\mathbf{a}}$. The estimate $\hat{\mathbf{C}}_{\mathbf{x}_k}$ of the covariance matrix of $\mathbf{x}_k$ is defined later in this section.

We select p_λ to be a conjugate pdf of λ, which is a complex Wishart pdf of dimension one [20, Sec. IVb]. From (6)

$$m_{p_\lambda\to\lambda} = p_{\mathcal{CW}_1}\left(M_{\mathrm{pr}}^{-1}, d_{\mathrm{pr}}\right) \quad (22)$$

with given parameters M_{pr} and d_{pr}. By inserting (21) and (22) into the message-passing rule (7), we obtain the complex Wishart pdf

$$m_{\lambda\to p_{\mathbf{y}}} \propto p_{\mathcal{CW}_1}\left((\hat{W}+M_{\mathrm{pr}})^{-1}, d_{\mathrm{pr}}+GNL\right) = b_\lambda. \quad (23)$$

It is enough to pass the first moment $\hat{\lambda}_{\mathrm{VMP}} = (d_{\mathrm{pr}}+GNL)(\hat{W}+M_{\mathrm{pr}})^{-1}$ [20, Eq. (22)] of this pdf, since the other message updates only depend on this value. As we have no prior information on λ, we select p_λ to be uniform over the range of λ. For this improper prior, we have $M_{\mathrm{pr}}=0$ and $d_{\mathrm{pr}}=0$ [20].

C. MIMO Decoding

To update $b_{\mathbf{x}_k}$, we compute the messages to and from the variable node $\mathbf{x}_k$. From (6), the message from node $p_{\mathbf{y}}$ to variable node $\mathbf{x}_k$ is

$$m_{p_{\mathbf{y}}\to\mathbf{x}_k} = \exp\left(\langle \ln\, p_{\mathbf{y}}(\mathbf{y}|\mathbf{a},\lambda,\mathbf{x})\rangle_{m_{\mathbf{a}\to p_{\mathbf{y}}}m_{\lambda\to p_{\mathbf{y}}}\prod_{k'\neq k} m_{\mathbf{x}_{k'}\to p_{\mathbf{y}}}}\right). \quad (24)$$

Channel Estimation

VMP receiver:

$$\hat{\mathbf{a}} = \left(\mathbf{C}_{\mathbf{a}}^{-1} + \hat{\lambda}_{\mathrm{VMP}}\hat{\mathbf{X}}^{\mathrm{H}}\hat{\mathbf{X}} + \hat{\lambda}_{\mathrm{VMP}}(\mathbf{I}_G \otimes \hat{\mathbf{C}}_{\mathbf{x}})\right)^{-1} \hat{\lambda}_{\mathrm{VMP}}\hat{\mathbf{X}}^{\mathrm{H}}\mathbf{y} \tag{9}$$

LMMSE-based receiver:

$$\hat{\mathbf{a}} = \left(\mathbf{C}_{\mathbf{a}}^{-1} + \hat{\mathbf{X}}^{\mathrm{H}}\hat{\mathbf{\Lambda}}^{\mathrm{chan}}_{\mathrm{LMMSE}}\hat{\mathbf{X}}\right)^{-1} \hat{\mathbf{X}}^{\mathrm{H}}\hat{\mathbf{\Lambda}}^{\mathrm{chan}}_{\mathrm{LMMSE}}\mathbf{y} \tag{10}$$

MIMO Detection/Interference Cancellation

VMP receiver:

$$\hat{\mathbf{x}}_k = \left(\hat{\lambda}_{\mathrm{VMP}}\hat{\mathbf{A}}_k^{\mathrm{H}}\hat{\mathbf{A}}_k + \hat{\lambda}_{\mathrm{VMP}}\sum_{g'}\sum_{g}\mathrm{Diag}(\mathrm{diag}(\hat{\mathbf{C}}_{\mathbf{a}_{kg}\mathbf{a}_{kg'}}))\right)^{-1} \hat{\lambda}_{\mathrm{VMP}}\hat{\mathbf{A}}_k^{\mathrm{H}}\left(\mathbf{y} - \sum_{k'\neq k}\hat{\mathbf{A}}_{k'}\hat{\mathbf{x}}_{k'}\right) \tag{11}$$

LMMSE-based receiver:

$$\hat{\mathbf{x}}_k = \left(\mathbf{C}_{\mathbf{x}_k}^{-1} + \hat{\mathbf{A}}_k^{\mathrm{H}}\hat{\mathbf{\Lambda}}^{\mathrm{det}}_{\mathrm{LMMSE}}\hat{\mathbf{A}}_k\right)^{-1} \hat{\mathbf{A}}_k^{\mathrm{H}}\hat{\mathbf{\Lambda}}^{\mathrm{det}}_{\mathrm{LMMSE}}\left(\mathbf{y} - \sum_{k'\neq k}\hat{\mathbf{A}}_{k'}\hat{\mathbf{x}}_{k'}\right) \tag{12}$$

Estimation of the Precision Matrix

VMP receiver: $\hat{\mathbf{\Lambda}}_{\mathrm{VMP}} = \hat{\lambda}_{\mathrm{VMP}}\mathbf{I}_{GNL}$ with

$$\hat{\lambda}_{\mathrm{VMP}} = \left(\frac{\mathrm{tr}\left((\mathbf{y}-\hat{\mathbf{A}}\hat{\mathbf{x}})(\mathbf{y}-\hat{\mathbf{A}}\hat{\mathbf{x}})^{\mathrm{H}} + \hat{\mathbf{X}}\hat{\mathbf{C}}_{\mathbf{a}}\hat{\mathbf{X}}^{\mathrm{H}} + \sum_k \hat{\mathbf{A}}_k\hat{\mathbf{C}}_{\mathbf{x}_k}\hat{\mathbf{A}}_k^{\mathrm{H}} + \sum_k(\mathbf{1}_G \otimes \hat{\mathbf{C}}_{\mathbf{x}_k}) \odot \mathrm{Diag}(\mathrm{diag}(\hat{\mathbf{C}}_{\mathbf{a}_k}))\right)}{GNL}\right)^{-1} \tag{13}$$

LMMSE-based receiver: σ_a^2 is the average power of the channel

$$\hat{\mathbf{\Lambda}}^{\mathrm{chan}}_{\mathrm{LMMSE}} = \left(\left(\frac{\mathrm{tr}\left(\left(\mathbf{y}-\hat{\mathbf{A}}\hat{\mathbf{x}}\right)\left(\mathbf{y}-\hat{\mathbf{A}}\hat{\mathbf{x}}\right)^{\mathrm{H}}\right)}{GNL}\right)\mathbf{I}_{GNL} + \sigma_a^2\mathbf{I}_G \otimes \sum_k \hat{\mathbf{C}}_{\mathbf{x}_k}\right)^{-1} \tag{14}$$

$$\hat{\mathbf{\Lambda}}^{\mathrm{det}}_{\mathrm{LMMSE}} = \left(\left(\frac{\mathrm{tr}\left(\left(\mathbf{y}-\hat{\mathbf{A}}\hat{\mathbf{x}}\right)\left(\mathbf{y}-\hat{\mathbf{A}}\hat{\mathbf{x}}\right)^{\mathrm{H}}\right)}{GNL}\right)\mathbf{I}_{GNL} + \sum_{k'\neq k}\hat{\mathbf{A}}_{k'}\hat{\mathbf{C}}_{\mathbf{x}_{k'}}\hat{\mathbf{A}}_{k'}^{\mathrm{H}}\right)^{-1} \tag{15}$$

Fig. 3. Channel estimation, MIMO detection/interference cancellation and the precision matrix estimation in the VMP and LMMSE-based receiver.

Solving the expectation, yields

$$m_{p_{\mathbf{y}}\to\mathbf{x}_k} \propto p_{\mathcal{CN}}(\hat{\mathbf{x}}_k, \hat{\mathbf{C}}_{\mathbf{x}_k}) \tag{25}$$

with mean vector $\hat{\mathbf{x}}_k = \hat{\lambda}_{\mathrm{VMP}}\hat{\mathbf{C}}_{\mathbf{x}_k}\hat{\mathbf{A}}_k^{\mathrm{H}}\big(\mathbf{y} - \sum_{k'\neq k}\hat{\mathbf{A}}_{k'}\hat{\mathbf{x}}_{k'}\big)$ and covariance matrix $\hat{\mathbf{C}}_{\mathbf{x}_k} = (\hat{\lambda}_{\mathrm{VMP}}\hat{\mathbf{A}}_k^{\mathrm{H}}\hat{\mathbf{A}}_k + \hat{\lambda}_{\mathrm{VMP}}\sum_{g'}\sum_{g}\mathrm{diag}(\hat{\mathbf{C}}_{\mathbf{a}_{kg}\mathbf{a}_{kg'}}))^{-1}$. The estimate $\hat{\mathbf{C}}_{\mathbf{a}_{kg}\mathbf{a}_{kg'}}$ of the cross-covariance matrix of the channel vectors $\mathbf{a}_{kg}$ and $\mathbf{a}_{kg'}$ can be obtained from $\hat{\mathbf{C}}_{\mathbf{a}}$.

Demodulation and decoding are performed in the right region of the graph in Fig. 2 using the SP algorithm. The estimated mean of a symbol x_{knl} in $\mathbf{x}_k$ is computed to be $\hat{x}_{knl} = \sum_{x\in\mathcal{M}} xP(x_{knl} = x|\hat{\mathbf{x}}_k)$, where $P(x_{knl} = x|\hat{\mathbf{x}}_k) = \sum_{\mathbf{x}_k\in\mathcal{X}_k, x_{knl}=x} m_{p_{\mathbf{y}}\to\mathbf{x}_k}(\mathbf{x}_k)$ with $\mathcal{M}$ denoting the set of constellation points of the selected M-QAM modulation. For convolutional codes, these marginals can be obtained with the BCJR algorithm. Likewise, the estimated variance of x_{knl} is $\hat{\sigma}^2_{x_{knl}} = \sum_{x\in\mathcal{M}} x^2P(x_{knl} = x|\hat{\mathbf{x}}_k) - \hat{x}^2_{knl}$. Any two distinct symbols are assumed to be uncorrelated. As a result, the estimate of the covariance matrix of $\mathbf{x}_k$ after decoding reads $\hat{\mathbf{C}}_{\mathbf{x}_k} = \mathrm{Diag}(\hat{\sigma}^2_{x_{k11}}, \ldots \hat{\sigma}^2_{x_{knL}})$.

We approximate the message from $\mathbf{x}_k$ to $p_{\mathbf{y}}$ by a Gaussian pdf. Notice that the Gaussian family is the conjugate family of $\mathbf{x}_k$ for $p_{\mathbf{y}}$. With this approximation and from (25) we obtain

$$m_{\mathbf{x}_k\to p_{\mathbf{y}}} \propto p_{\mathcal{CN}}(\hat{\mathbf{x}}_k, \hat{\mathbf{C}}_{\mathbf{x}_k}) = b_{\mathbf{x}_k}. \tag{26}$$

We only pass the natural statistics $\hat{\mathbf{x}}_k$, $\hat{\mathbf{C}}_{\mathbf{x}_k}$ to $p_{\mathbf{y}}$. From (8), the message (26) represents the estimated posterior pdf of $\mathbf{x}_k$.

V. Comparison with the LMMSE-based receiver

In this section, we compare the VMP receiver derived in the previous section to a state-of-the-art iterative receiver proposed in [5], further developed for detection in multiuser CDMA [6], and applied to MIMO-OFDM systems in [7]. We refer to this receiver as the LMMSE-based receiver. Due to lack of space, the derivation of the LMMSE-based receiver is not included in this work, but the expressions of the different component blocks are summarized in Fig. 3 together with the corresponding expressions obtained for the VMP receiver.

The conceptual difference between the two schemes is that in the LMMSE-based receiver the different constituent blocks are designed independently, while in the VMP receiver the blocks corresponding to factors in the VMP region are designed jointly, by minimizing a global cost function, i.e. a KL divergence, in this region.

By inspecting the expressions in Fig. 3 we observe that the LMMSE-based receiver and the VMP receiver share some structural properties. For instance, from (9) and (10) it is clear that both algorithms use an LMMSE-like channel estimator, which mainly depends on the channel prior covariance, estimates of the transmitted symbols and an estimate of the precision matrix, namely $\hat{\lambda}_{\mathrm{VMP}}\mathbf{I}$ in the VMP receiver and $\hat{\mathbf{\Lambda}}^{\mathrm{chan}}_{\mathrm{LMMSE}}$ in the LMMSE-based receiver. Similarly, the detection part of both receivers consists of interference cancellation followed by LMMSE filtering of the residual interference. However, we can highlight two critical differences between

978-1-4244-5638-3/10/$26.00 ©2010 IEEE

TABLE I
PARAMETER SETTINGS FOR THE SIMULATIONS

Cyclic prefix length	4.7 μs
Symbol duration	66.7 μs
Subcarrier spacing	15 kHz
Pilot overhead	4.8% pilots
Pilot pattern	Regular spacing/diamond, QPSK
Modulation alphabet	16-QAM
Number of information bits	660
Number of subcarriers N	75
Number of OFDM symbols L	7
Number of transmitters K	2
Number of receivers G	2
Channel interleaver	block
Convolutional code	$(155, 117, 127)_8$

the two algorithms: firstly, only one scalar estimate of the precision parameter is needed in the VMP receiver, while the LMMSE-based receiver calculates two different precision matrices, one for channel estimation ($\hat{\mathbf{\Lambda}}^{\text{chan}}_{\text{LMMSE}}$) and one for detection ($\hat{\mathbf{\Lambda}}^{\text{det}}_{\text{LMMSE}}$); secondly, the LMMSE-based receiver does not deal with the uncertainty in the channel weight estimates and considers them as the true values in the detection part, while the VMP receiver accounts for channel estimation errors via the term $\hat{\mathbf{C}}_{\mathbf{a}}$ in (11) and (13).

VI. Simulation Results

To verify the performance of the VMP receiver, we perform Monte-Carlo simulations for an LTE-like 2×2 system with the settings reported in Table I. We consider a pilot scheme where all transmitters transmit pilots in the same time-frequency resources. Realizations of the channel time-frequency response are generated using the extended typical urban (ETU) channel model from the 3GPP LTE standard [21], with Rayleigh-fading channel taps, and assuming no correlation over transmit or receive antennas. Note that the channel is wide-sense-stationary and uncorrelated-scattering (WSSUS) [22]. We compute the prior covariance matrix $\mathbf{C}_{\mathbf{a}}$ from the channel time-frequency correlation function.

We test the OFDM-MIMO system with the two receivers described in Fig. 3. Both receivers use the same initialization, consisting of MMSE pilot-based channel estimation and joint soft-decision maximum likelihood (ML) detection, followed by soft-in soft-out sequential decoding. In both receivers an iteration consists of estimation of the channel weights, followed by sequential detection and decoding of all K transmitted frames, and ending with estimation of the precision parameter or matrices.

The bit-error-rate (BER) performance of both receivers versus the signal-to-noise ratio (E_b/N_0) is illustrated in Fig. 4. For the sake of comparison, the initialization is also depicted (denoted by the 'Linear Receiver' tag). Both receivers perform 10 iterations. The results show that both iterative structures significantly improve the performance of the linear receiver, especially for E_b/N_0 larger than 0 dB. Moreover, the VMP receiver outperforms the LMMSE-based receiver in the considered signal-to-noise range. The gain is about 0.5 dB in the operation range of the MIMO-OFDM system. The convergence behaviour of both iterative structures is described

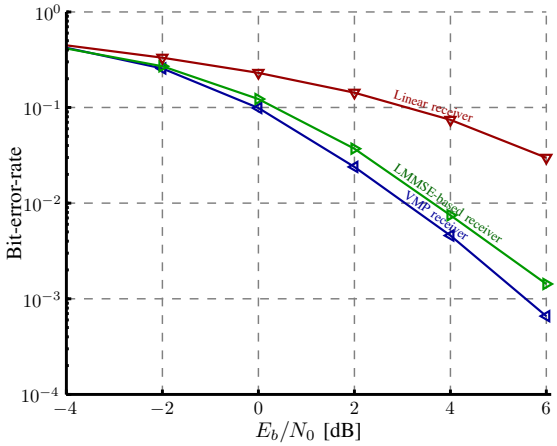

Fig. 4. Coded bit-error-rate.

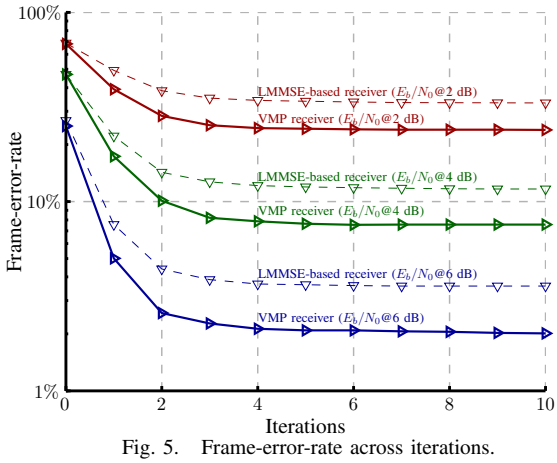

Fig. 5. Frame-error-rate across iterations.

in Fig. 5, which depicts the frame-error-rate versus the number of iterations at the receiver for three different E_b/N_0 values. Both receivers converge after approximately 5 iterations for all operation points. Again, the VMP receiver outperforms the LMMSE-based receiver regardless of the number of iterations.

VII. Conclusion

We derive a novel iterative receiver structure for M-QAM MIMO-OFDM operating in frequency-selective time-variant channels. The scheme performs jointly semi-blind estimation of the channel weights and of the noise inverse variance based on both data and pilot symbols, serial interference cancellation, and decoding. The scheme was already proposed for CDMA in [11]. A variational message-passing (VMP) interpretation of it is provided here.

The VMP receiver is compared with the LMMSE-based iterative receiver derived in [5]–[7]. Both iterative architectures are made of the same blocks and exhibit similar complexity. However, in the VMP receiver all blocks but decoding are jointly optimized according to a global cost function, the KL divergence, while in the LMMSE-based receiver all blocks are designed independently. Furthermore, the VMP framework yields a structure that takes into account the inaccuracy of the channel weight estimates. This inaccuracy is neglected in the LMMSE-based receiver.

In order to assess the effect of these structural differences,

978-1-4244-5638-3/10/$26.00 ©2010 IEEE

we evaluate the performance of both receivers in an LTE-like scenario. The simulation results show that the VMP receiver outperforms the LMMSE-based receiver with a signal-to-noise ratio gain of 0.5 dB at relevant BER values.

An issue not addressed in the paper is how to combine efficiently the VMP algorithm – used for channel weight and noise inverse variance estimation as well as serial interference cancellation – and the sum-product algorithm – employed for decoding – in the receiver. A solution has been recently proposed in [23].

VIII. Acknowledgements

The authors would like to thank NOKIA Denmark for the financial support which made this work possible. This work has also been supported in part by the 4GMCT cooperative research project funded by Infineon Technologies Denmark A/S, Agilent Technologies, Aalborg University and the Danish National Advanced Technology Foundation, by the European Commission within the FP7-ICT Network of Excellence in Wireless Communications, NEWCOM++ (Contract No. 216715) and by WWTF grant ICT08-44, FWF grant S10603-N13 within the National Research Network SISE.

Appendix A
The Variational Message-Passing Algorithm

In VB inference [11] we consider as the cost function the KL divergence $\mathcal{D}_{\mathrm{KL}}(b_{\mathbf{\Phi}} \| f_{\mathbf{\Phi}}) \triangleq \int \mathrm{d}\mathbf{\Phi}\; b_{\mathbf{\Phi}} \log \frac{b_{\mathbf{\Phi}}}{f_{\mathbf{\Phi}}}$, where $f_{\mathbf{\Phi}}$ is a pdf of a set of variables $\mathbf{\Phi}$ and $b_{\mathbf{\Phi}}$ is an auxiliary function, which approximates $f_{\mathbf{\Phi}}$. We seek an auxiliary function that minimizes the cost function.

We reformulate the VB inference problem to message-passing on a factor-graph [16]. We assume that $f_{\mathbf{\Phi}}$ factorizes according to $f_{\mathbf{\Phi}} = \prod_a f_a(\mathcal{N}_{f_a})$, where $\mathcal{N}_{f_a} \subseteq \mathbf{\Phi}$ is the set of neighbouring variables of f_a. We select an auxiliary function $b_{\mathbf{\Phi}}$, which factorizes according to $b_{\mathbf{\Phi}} = \prod_{\phi \in \mathbf{\Phi}} b_\phi$. As shown in [11], the factor b_ϕ of $b_{\mathbf{\Phi}}$ which minimizes $\mathcal{D}_{\mathrm{KL}}(b_{\mathbf{\Phi}} \| f_{\mathbf{\Phi}})$ with all other factors $b_{\phi'}$, $\forall \phi' \in \mathbf{\Phi} \backslash \phi$ fixed is

$$b_\phi \propto \exp\langle \sum_{f_a \in \mathcal{N}_\phi} \ln f_a \rangle_{b_{\phi'} \forall \phi' \in \mathcal{N}_{f_a} \backslash \phi} \tag{27}$$

$$\propto \prod_{f_a \in \mathcal{N}_\phi} \exp\langle \ln f_a \rangle_{b_{\phi'} \forall \phi' \in \mathcal{N}_{f_a} \backslash \phi}, \tag{28}$$

where $\mathcal{N}_\phi$ is the set of neighbouring factors of ϕ. With the definitions in (6) and (7) we can recast (28) as

$$b_\phi \propto \prod_{f'_a \in \mathcal{N}_\phi} m_{f'_a \to \phi} = m_{\phi \to f_a} \tag{29}$$

for any $f_a \in \mathcal{N}_\phi$, and we have

$$b_{\mathbf{\Phi}} = \prod_{\phi \in \mathbf{\Phi}} b_\phi \propto \prod_{\phi \in \mathbf{\Phi}} \prod_{f'_a \in \mathcal{N}_\phi} m_{f'_a \to \phi}. \tag{30}$$

Identity (28) can be used to design an iterative algorithm which at each iteration updates a given factor b_ϕ of $b_{\mathbf{\Phi}}$ while keeping the other factors fixed. The iterative algorithm converges in the sense of the KL divergence, since $\mathcal{D}_{\mathrm{KL}}(b_{\mathbf{\Phi}} \| f_{\mathbf{\Phi}})$ is minimized at each iteration. The identities in (29) provide a message-passing interpretation of the updating steps.

References

[1] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo codes," in *Proc. IEEE Int. Conf. Commun. (ICC'93)*, May 1993, pp. 1064–1070.

[2] R. Koetter, A. C. Singer, and M. Tücher, "Turbo equalization," *IEEE Trans. Signal Processing*, vol. 1, pp. 64–80, 2004.

[3] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithms," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[4] X. Wang and H. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, pp. 1046–1061, Jul. 1999.

[5] M. Lončar, R. Müller, J. Wehinger, C. Mecklenbräuker, and T. Abe, "Iterative channel estimation and data detection in frequency-selective fading MIMO channels," *Euro. T. Telecom.*, vol. 15, pp. 459–470, 2004.

[6] J. Wehinger and C. Mecklenbräuker, "Iterative CDMA multiuser receiver with soft decision-directed channel estimation," *IEEE Trans. Signal Processing*, vol. 54, pp. 3922–3934, Oct. 2006.

[7] P. Rossi and R. Müller, "Joint twofold-iterative channel estimation and multiuser detection for MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 11, pp. 4719–4729, Nov. 2008.

[8] H. Attias, "A variational Bayesian framework for graphical models," *Advances in Neural Information Processing Systems*, pp. 209–215, 2000.

[9] M. Beal, "Variational algorithms for approximate inference," Ph.D. dissertation, University of Cambridge, May 2003.

[10] L. Christensen and J. Larsen, "On data and parameter estimation using the variational Bayesian EM-algorithm for block-fading frequency-selective MIMO channels," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Sign. Process. (ICASSP'06)*, vol. 4, 2006, pp. 465–468.

[11] B. Hu, I. Land, L. Rasmussen, R. Piton, and B. Fleury, "A divergence minimization approach to joint multiuser decoding for coded CDMA," *IEEE J. Select. Areas Commun.*, vol. 26, no. 3, pp. 432–445, Apr. 2008.

[12] D. Lin and T. Lim, "The variational inference approach to joint data detection and phase noise estimation in OFDM," *IEEE Trans. Signal Processing*, vol. 55, no. 5, pp. 1862–1874, May 2007.

[13] M. Nissilä, "Iterative receivers for digital communications via variational inference and estimation," Ph.D. dissertation, Oulu University, 2008.

[14] X.-Y. Zhang, D.-G. Wang, and J.-B. Weiss, "Joint symbol detection and channel estimation for MIMO-OFDM systems via the variational Bayesian EM algorithms," in *Proc. IEEE Wireless Commun. and Networking Conf. (WCNC'08)*, Mar.-Apr. 2008, pp. 13–17.

[15] C. N. Manchon, G. Kirkelund, B. Fleury, P. Mogensen, L. Deneire, T. Sorensen, and C. Rom, "Interference cancellation based on divergence minimization for MIMO-OFDM receivers," in *Proc. IEEE Global Telecom. Conf. (GLOBECOM'09)*, Dec. 2009.

[16] J. Dauwels, "On variational message passing on factor graphs," in *IEEE Int. Symp. on Inform. Theory (ISIT'07)*, Jun. 2007, pp. 2546–2550.

[17] J. Winn and C. Bishop, "Variational message passing," *Journal of Machine Learning Research*, 2004.

[18] H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. Kschischang, "The factor graph approach to model-based signal processing," *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1295–1322, Jun. 2007.

[19] J. Tauge and C. Caldwell, "Expectations of useful complex Wishart forms," *Multidimensional Systems and Signal Processing Archive*, vol. 5, pp. 263–278, 1994.

[20] L. Svensson and M. Lundberg, "On posterior distributions for signals in Gaussian noise with unknown covariance matrix," *IEEE Trans. Signal Processing*, vol. 53, no. 9, pp. 3554–3571, Sept. 2005.

[21] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); LTE Physical Layer - base station (BS) radion transmission and reception (release 8)," 3GPP, Tech. Rep. TS 36.104, V8.8.0, Dec. 2009.

[22] P. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun.*, vol. 11, no. 4, pp. 360–393, Dec. 1963.

[23] E. Riegler, G. E. Kirkelund, C. N. Manchón, and B. H. Fleury, "Merging belief propagation and the mean field approximation: a free energy approach," in *Proc. 6th Int. Symp. on Turbo Codes & Iterative Information Processing, (ISTC'10)*, accepted for publication.