Aalborg Universitet

AALBORG UNIVERSITY

Message-Passing Algorithms for Channel Estimation and Decoding Using Approximate Inference

Badiu, Mihai Alin; Kirkelund, Gunvor Elisabeth; Manchón, Carles Navarro; Riegler, Erwin; Fleury, Bernard Henri

Message-Passing Algorithms for Channel Estimation and Decoding Using Approximate Inference

Mihai-A. Badiu*†, Gunvor E. Kirkelund*, Carles Navarro Manchón*, Erwin Riegler‡ and Bernard H. Fleury*

* Aalborg University, Denmark
† Technical University of Cluj-Napoca, Romania
‡ Vienna University of Technology, Austria

Abstract—We design iterative receiver schemes for a generic communication system by treating channel estimation and information decoding as an inference problem in graphical models. We introduce a recently proposed inference framework that combines belief propagation (BP) and the mean field (MF) approximation and includes these algorithms as special cases. We also show that the expectation propagation and expectation maximization (EM) algorithms can be embedded in the BP-MF framework with slight modifications. By applying the considered inference algorithms to our probabilistic model, we derive four different message-passing receiver schemes. Our numerical evaluation in a wireless scenario demonstrates that the receiver based on the BP-MF framework and its variant based on BP-EM yield the best compromise between performance, computational complexity and numerical stability among all candidate algorithms.

I. INTRODUCTION

The design of advanced receiver algorithms is crucial to meet the stringent requirements of modern communication systems. Motivated by the successful application of the "turbo" principle in the decoding of channel codes, a large number of works have been devoted to the design of turbo receivers (see [1] and the references therein). While in many of these works the receiver modules are individually designed and heuristically interconnected to exchange soft values, iterative receiver algorithms can be rigourously designed and better understood as instances of message-passing inference techniques (e.g., see [2]).

In this context, variational Bayesian inference in probabilistic models [3] have proven to be a very useful tool to design receivers where tasks like channel estimation, detection and decoding are jointly derived. Among the variational techniques, belief propagation (BP) [4], [5] has found the most widespread use. Originally applied to the decoding of channel codes, BP has been shown to be especially efficient in discrete probabilistic models. An alternative to BP is the mean field (MF) approximation and its message-passing counterpart, usually referred to as variational message-passing [6]. MF inference has been successfully applied to continuous probabilistic models involving probability density functions (pdfs) belonging to an exponential family, in which BP suffers from numerical intractability. Other notable examples of general-purpose inference techniques are expectation-maximization (EM) [7] and expectation propagation (EP) [8]. EM is a special case of MF, where the approximate pdfs – referred to as *beliefs* – are Dirac delta functions; EP can be seen as an approximation of BP where some beliefs are approximated by pdfs in a specific exponential family. Some attempts to find a unified framework encompassing all these techniques include the α-divergence interpretation in [9] and the region-based free energy approximations in [10]. Following

the latter approach, a novel hybrid message-passing inference framework combining BP and the MF approximation was recently proposed in [11].

In this paper, we investigate the design of receivers that perform joint channel estimation and data decoding in a generic communication system. For this purpose, we capitalize on the combined inference framework [11], which provides some degree of freedom in the choice of the parts of the factor graph in which either BP or MF is applied. We show that this framework can be modified to naturally embed EP, EM and BP with Gaussian approximation of some messages. Then, we apply these hybrid inference techniques to the underlying probabilistic model of the system and obtain four receiver algorithms, whose performance we assess by simulating a wireless system.

Notation: we denote by $|\mathcal{I}|$ the cardinality of a finite set $\mathcal{I}$; the relative complement of $\{i\}$ in $\mathcal{I}$ is written as $\mathcal{I} \setminus i$; the set $\{i \in \mathbb{N} \mid 1 \le i \le n\}$ is denoted by $[1:n]$. Boldface lowercase and uppercase letters are used to represent vectors and matrices, respectively; superscripts $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ denote transposition and Hermitian transposition, respectively. The Hadamard product of two vectors is denoted by $\odot$. For a vector $\mathbf{x} = (x_i \mid i \in \mathcal{I})^{\mathrm{T}}$, we write $\mathbf{x}_{\bar{i}} = (x_j \mid j \in \mathcal{I} \setminus i)^{\mathrm{T}}$; for a matrix $\mathbf{A} \in \mathbb{C}^{m \times n}$, $[\mathbf{A}]_{i,j}$ denotes its (i,j)th entry, $[\mathbf{A}]_{\bar{i},\bar{j}}$ is the matrix $\mathbf{A}$ with the ith row and jth column deleted, $[\mathbf{A}]_{\bar{i},j}$ denotes the column vector $([\mathbf{A}]_{k,j} \mid k \in [1:m] \setminus i)^{\mathrm{T}}$, and $[\mathbf{A}]_{i,\bar{j}}$ is the row vector $([\mathbf{A}]_{i,k} \mid k \in [1:n] \setminus j)$. The pdf of a multivariate complex Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted by $\mathrm{CN}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$. We write $f(x) \propto g(x)$ when $f(x) = cg(x)$ for some positive constant c. We denote by $\mathcal{G}[\cdot]$ the approximation of the pdf in the argument with a Gaussian pdf with the same mean and covariance matrix. The Dirac delta function is denoted by $\delta(\cdot)$.

II. MESSAGE-PASSING INFERENCE ALGORITHMS

We begin by concisely describing the unified message-passing algorithm that combines the BP and MF approaches (refer to [11]). Then, we briefly show how other widespread inference algorithms can be obtained as particular instances or slight modifications of the unified framework.

Let $p(\mathbf{z})$ be an arbitrary pdf of a random vector $\mathbf{z} \triangleq (z_i \mid i \in \mathcal{I})^{\mathrm{T}}$ which factorizes as

$$p(\mathbf{z}) = \prod_{a \in \mathcal{A}} f_a(\mathbf{z}_a) = \prod_{a \in \mathcal{A}_{\mathrm{MF}}} f_a(\mathbf{z}_a) \prod_{c \in \mathcal{A}_{\mathrm{BP}}} f_c(\mathbf{z}_c) \quad (1)$$

where $\mathbf{z}_a$ is the vector of all variables z_i that are arguments of the function f_a. We have grouped the factors into two sets that partition $\mathcal{A}$: $\mathcal{A}_{\mathrm{MF}} \cap \mathcal{A}_{\mathrm{BP}} = \emptyset$ and $A_{\mathrm{MF}} \cup \mathcal{A}_{\mathrm{BP}} = \mathcal{A}$. The

factorization in (1) can be visualized in a factor graph [4] representation. We define $\mathcal{N}(a) \subseteq \mathcal{I}$ to be the set of indices of all variables z_i that are arguments of function f_a; similarly, $\mathcal{N}(i) \subseteq \mathcal{A}$ denotes the set of indices of all functions f_a that depend on z_i. The parts of the graph that correspond to $\prod_{a \in \mathcal{A}_{\mathrm{BP}}} f_a(\mathbf{z}_a)$ and to $\prod_{a \in \mathcal{A}_{\mathrm{MF}}} f_a(\mathbf{z}_a)$ are referred to as "BP part" and "MF part", respectively. We denote the variable nodes in the BP part by $\mathcal{I}_{\mathrm{BP}} \triangleq \bigcup_{a \in \mathcal{A}_{\mathrm{BP}}} \mathcal{N}(a)$ and those in the MF part by $\mathcal{I}_{\mathrm{MF}} \triangleq \bigcup_{a \in \mathcal{A}_{\mathrm{MF}}} \mathcal{N}(a)$.

The combined BP-MF inference algorithm approximates the marginals $p(z_i) = \int p(\mathbf{z}) \mathrm{d} \mathbf{z}_{\bar{i}}$, $i \in \mathcal{I}$ by auxiliary pdfs $b_i(z_i)$ called *beliefs*. They are computed as [11]

$$b_i(z_i) = \omega_i \prod_{a \in \mathcal{A}_{\mathrm{BP}} \cap \mathcal{N}(i)} m_{a \to i}^{\mathrm{BP}}(z_i) \prod_{a \in \mathcal{A}_{\mathrm{MF}} \cap \mathcal{N}(i)} m_{a \to i}^{\mathrm{MF}}(z_i) \quad (2)$$

with

$$m_{a \to i}^{\mathrm{BP}}(z_i) = \omega_a \int \prod_{j \in \mathcal{N}(a) \setminus i} \mathrm{d} z_j \, n_{j \to a}(z_j) \, f_a(\mathbf{z}_a),$$
$$\forall \, a \in \mathcal{A}_{\mathrm{BP}}, i \in \mathcal{N}(a)$$

$$m_{a \to i}^{\mathrm{MF}}(z_i) = \exp \left(\int \prod_{j \in \mathcal{N}(a) \setminus i} \mathrm{d} z_j \, n_{j \to a}(z_j) \ln f_a(\mathbf{z}_a) \right),$$
$$\forall \, a \in \mathcal{A}_{\mathrm{MF}}, i \in \mathcal{N}(a)$$

$$n_{i \to a}(z_i) = \omega_i \prod_{c \in \mathcal{A}_{\mathrm{BP}} \cap \mathcal{N}(i) \setminus a} m_{c \to i}^{\mathrm{BP}}(z_i) \prod_{c \in \mathcal{A}_{\mathrm{MF}} \cap \mathcal{N}(i)} m_{c \to i}^{\mathrm{MF}}(z_i),$$
$$\forall \, i \in \mathcal{N}(a), a \in \mathcal{A},$$
$$(3)$$

where ω_i and ω_a are constants that ensure normalized beliefs.

Belief propagation is obtained as a particular case of BP-MF by setting $\mathcal{A}_{\mathrm{MF}} = \emptyset$, since in this case the expressions in (3) reduce to the BP message computations. Similarly, *mean field* is an instance of BP-MF when $\mathcal{A}_{\mathrm{BP}} = \emptyset$.

Expectation propagation is very similar to BP, the main difference being that it constrains the beliefs of some variables to be members of a specific exponential family. Assuming Gaussian approximations of the beliefs, EP can also be integrated in the BP-MF framework by modifying the messages

$$m_{a \to i}^{\mathrm{EP}}(z_i) \propto \frac{1}{n_{i \to a}(z_i)} \, \mathcal{G} \left[n_{i \to a}(z_i) \, m_{a \to i}^{\mathrm{BP}}(z_i) \right], \quad (4)$$

for all $i \in \mathcal{I}_{\mathrm{EP}} \subseteq \mathcal{I}_{\mathrm{BP}}$, $a \in \mathcal{N}(i) \cap \mathcal{A}_{\mathrm{BP}}$.

The *expectation-maximization* algorithm is a special case of MF when the beliefs of some variables are constrained to be Dirac delta functions [11]. Again, we include this approximation in the BP-MF framework. This leads to $n_{i \to a}(z_i) = \delta(z_i - \tilde{z}_i)$ for all $i \in \mathcal{I}_{\mathrm{EM}} \subseteq \mathcal{I}_{\mathrm{MF}}$ and $a \in \mathcal{N}(i) \cap \mathcal{A}_{\mathrm{MF}}$, where $\tilde{z}_i$ maximizes the unconstrained belief (2). We refer to this modified algorithm as BP-EM.

III. Probabilistic System Model

In this section, we present the signal model of our inference problem and its graphical representation. These will establish the baseline for the derivation of message-passing receivers.

We analyze a system consisting of one transmitter and one receiver. A message represented by a vector $\mathbf{u} = (u_k \mid k \in [1 : K])^{\mathrm{T}} \in \{0, 1\}^K$ of information bits is conveyed

by sending N data and M pilot channel symbols having the sets of indices $\mathcal{D} \subseteq [1 : M + N]$ and $\mathcal{P} \subseteq [1 : M + N]$, respectively, such that $\mathcal{D} \cup \mathcal{P} = [1 : M + N]$ and $\mathcal{D} \cap \mathcal{P} = \emptyset$. Specifically, vector $\mathbf{u}$ is encoded and interleaved using a rate $R = K/(NL)$ channel code and a random interleaver into the vector $\mathbf{c} = (\mathbf{c}_n^{\mathrm{T}} \mid \mathbf{c}_n \in \{0, 1\}^L, n \in [1 : N])^{\mathrm{T}}$ of length NL. For each $n \in [1 : N]$, the subvector $\mathbf{c}_n = (c_n^{(1)}, \dots, c_n^{(L)})^{\mathrm{T}}$ is mapped to a data symbol $x_{i_n} \in \mathcal{S}_{\mathrm{D}}$ with $i_n \in \mathcal{D}$, where $\mathcal{S}_{\mathrm{D}}$ is a discrete complex modulation alphabet of size 2^L. Symbols $\mathbf{x}_{\mathrm{D}} = (x_i \mid i \in \mathcal{D})^{\mathrm{T}}$ are multiplexed with pilot symbols $\mathbf{x}_{\mathrm{P}} = (x_j \mid j \in \mathcal{P})^{\mathrm{T}}$, which are randomly selected from a QPSK modulation alphabet. Finally, the aggregate vector of channel symbols $\mathbf{x} = (x_i \mid i \in \mathcal{D} \cup \mathcal{P})^{\mathrm{T}}$ is sent through a channel with the following input-output relationship:

$$\mathbf{y} = \mathbf{h} \odot \mathbf{x} + \mathbf{w}. \quad (5)$$

The vector $\mathbf{y} = (y_i \mid i \in [1 : M + N])^{\mathrm{T}}$ contains the received signal samples, $\mathbf{h} = (h_i \mid i \in [1 : M + N])^{\mathrm{T}}$ is the vector of channel coefficients, and $\mathbf{w} = (w_i \mid i \in [1 : M + N])^{\mathrm{T}}$ contains the samples of additive noise and has the pdf $p(\mathbf{w}) = \mathrm{CN}(\mathbf{w}; \mathbf{0}, \gamma^{-1} \mathbf{I}_{M+N})$ for some positive component precision γ. Note that (5) can model any channel with a multiplicative effect that is not affected by inter-symbol interference, e.g., a time-varying frequency-flat channel or the equivalent channel in the frequency domain in a multicarrier system.

Based on the above signal model, we can state the probabilistic model which captures the dependencies between the system variables. The pdf of the collection of observed and unknown variables factorizes as

$$p(\mathbf{y}, \mathbf{h}, \mathbf{x}_{\mathrm{D}}, \mathbf{c}, \mathbf{u}) = f_{\mathrm{H}}(\mathbf{h}) \prod_{i \in \mathcal{D}} f_{\mathrm{D}_i}(h_i, x_i) \prod_{j \in \mathcal{P}} f_{\mathrm{P}_j}(h_j)$$
$$\times \prod_{n \in [1:N]} f_{\mathrm{M}_n}(x_{i_n}, \mathbf{c}_n) f_{\mathrm{C}}(\mathbf{c}, \mathbf{u}) \prod_{k \in [1:K]} f_{\mathrm{U}_k}(u_k), \quad (6)$$

where $f_{\mathrm{D}_i}(h_i, x_i) \triangleq p(y_i | h_i, x_i)$ and $f_{\mathrm{P}_j}(h_j) \triangleq p(y_j | h_j)$ incorporate the observations in $\mathbf{y}$ and are given by

$$f_{\mathrm{D}_i}(h_i, x_i) = \mathrm{CN}\left(h_i x_i; y_i, \gamma^{-1}\right), \quad \forall i \in \mathcal{D}, \quad (7)$$
$$f_{\mathrm{P}_j}(h_j) = \mathrm{CN}\left(h_j x_j; y_j, \gamma^{-1}\right), \quad \forall j \in \mathcal{P}, \quad (8)$$

$f_{\mathrm{H}}(\mathbf{h}) \triangleq p(\mathbf{h})$ is the prior pdf of the vector of channel coefficients for which we set

$$f_{\mathrm{H}}(\mathbf{h}) = \mathrm{CN}\left(\mathbf{h}; \boldsymbol{\mu}_{\mathbf{h}}^{\mathrm{p}}, \boldsymbol{\Sigma}_{\mathbf{h}}^{\mathrm{p}}\right), \quad (9)$$

$f_{\mathrm{M}_n}(x_{i_n}, \mathbf{c}_n) \triangleq p(x_{i_n} | \mathbf{c}_n)$ stand for the modulation mapping, $f_{\mathrm{C}}(\mathbf{c}, \mathbf{u}) \triangleq p(\mathbf{c} | \mathbf{u})$ accounts for the coding and interleaving operations and $f_{\mathrm{U}_k}(u_k) \triangleq p(u_k)$ is the prior pmf of the kth information bit. To obtain (6), we used the fact that $\mathbf{y}$ is conditionally independent of $\mathbf{c}$ and $\mathbf{u}$ given $\mathbf{x}_{\mathrm{D}}$, $\mathbf{h}$ is independent of $\mathbf{x}_{\mathrm{D}}$, $\mathbf{c}$ and $\mathbf{u}$, the noise samples w_i are i.i.d., and each data symbol x_{i_n} is conditionally independent of all the other symbols given $\mathbf{c}_n$. The factorization in (6) can be visualized in the factor graph depicted in Fig. 1. The graph of the code and interleaver is not explicitly given, its structure being captured by f_{C}.

IV. Message-Passing Receiver Schemes

In this section, we derive iterative receiver schemes by applying different inference algorithms to the factor graph in
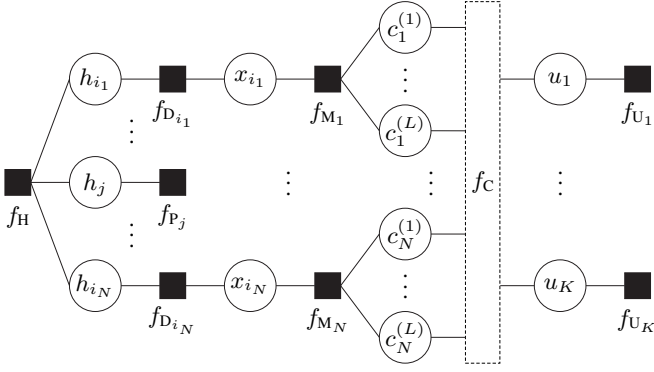
Fig. 1. Factor graph representation of the pdf factorization in (6) with $i_1, \ldots, i_N \in \mathcal{D}$ and $j \in \mathcal{P}$.

Fig. 1. The receiver has to infer the beliefs of the information bits using the observed vector $\mathbf{y}$ and prior knowledge, i.e., the pilot symbols and their set of indices $\mathcal{P}$, the noise precision γ, the channel statistics in (9), the modulation mapping and the structure of the channel code and interleaver.

We set $\mathcal{A}$ and $\mathcal{I}$ (defined in Section II for a general probabilistic model) to be the sets of all factors and variables, respectively, contained in our probabilistic model. Next, we show that the BP algorithm resulting from setting $\mathcal{A}_{\mathrm{MF}} = \emptyset$ yields messages of an intractable complexity. Assume that by running BP in the part of the graph containing the modulation and code constraints we obtain the messages

$$m_{f_{\mathrm{M}_n} \to x_{i_n}}^{\mathrm{BP}}(x_{i_n}) \propto \sum_{s \in \mathcal{S}_{\mathrm{D}}} \beta_{i_n}(s) \delta(x_{i_n} - s), \quad (10)$$

with $i_n \in \mathcal{D}, \forall n \in [1:N]$, where $\beta_{i_n}(s)$ represent extrinsic information on symbol x_{i_n}. These messages are further passed as $n_{x_{i_n} \to f_{\mathrm{D}_{i_n}}}(x_{i_n}) = m_{f_{\mathrm{M}_n} \to x_{i_n}}^{\mathrm{BP}}(x_{i_n})$. Then, for each $i \in \mathcal{D}$, compute the message

$$m_{f_{\mathrm{D}_i} \to h_i}^{\mathrm{BP}}(h_i) \propto \int f_{\mathrm{D}_i}(h_i, x_i) n_{x_i \to f_{\mathrm{D}_i}}(x_i) \, \mathrm{d}x_i$$
$$\propto \sum_{s \in \mathcal{S}_{\mathrm{D}}} \frac{\beta_i(s)}{|s|^2} \mathrm{CN} \left(h_i; \frac{y_i s^*}{|s|^2}, \frac{1}{\gamma |s|^2} \right), \quad (11)$$

while for all $j \in \mathcal{P}$ set

$$m_{f_{\mathrm{P}_j} \to h_j}^{\mathrm{BP}}(h_j) \propto f_{\mathrm{P}_j}(h_j) \propto \mathrm{CN} \left(h_j; \frac{y_j x_j^*}{|x_j|^2}, \frac{1}{\gamma |x_j|^2} \right). \quad (12)$$

Note that the message in (11) is proportional to a mixture of Gaussian pdfs with $|\mathcal{S}_{\mathrm{D}}| = 2^L$ components. Then, after setting $n_{h_i \to f_{\mathrm{H}}}(h_i) = m_{f_{\mathrm{D}_i} \to h_i}^{\mathrm{BP}}(h_i)$ for all $i \in \mathcal{D}$ and $n_{h_i \to f_{\mathrm{H}}}(h_i) = m_{f_{\mathrm{P}_i} \to h_i}^{\mathrm{BP}}(h_i)$ for all $i \in \mathcal{P}$, the message from f_{H} to h_i reads

$$m_{f_{\mathrm{H}} \to h_i}^{\mathrm{BP}}(h_i) \propto \int f_{\mathrm{H}}(\mathbf{h}) \prod_{j \in (D \cup P) \setminus i} n_{h_j \to f_{\mathrm{H}}}(h_j) \, \mathrm{d}h_j. \quad (13)$$

Using (9), (11) and (12), the message in (13) becomes a Gaussian mixture with $2^{L(N-1)}$ and 2^{LN} components for $i \in \mathcal{D}$ and $i \in \mathcal{P}$, respectively. Clearly, the computation of such messages is intractable and one has to use approximations.

A. Algorithm based on BP combined with Gaussian approximation

Since the intractability of the messages occurs due to the Gaussian mixture in (11), we approximate those messages as

proposed in [12], i.e., for each $i \in \mathcal{D}$ we set

$$m_{f_{\mathrm{D}_i} \to h_i}^{\mathrm{BP-GA}}(h_i) \propto \mathcal{G} \left[m_{f_{\mathrm{D}_i} \to h_i}^{\mathrm{BP}}(h_i) \right] = \mathrm{CN} \left(h_i; \mu_{h_i,\mathrm{o}}, \sigma_{h_i,\mathrm{o}}^2 \right) \quad (14)$$

with

$$\mu_{h_i,\mathrm{o}} = \sum_{s \in \mathcal{S}_{\mathrm{D}}} \alpha_i(s) \frac{y_i s^*}{|s|^2},$$
$$\sigma_{h_i,\mathrm{o}}^2 = \sum_{s \in \mathcal{S}_{\mathrm{D}}} \alpha_i(s) \left(\frac{|y_i|^2}{|s|^2} + \frac{1}{\gamma |s|^2} \right) - |\mu_{h_i,\mathrm{o}}|^2. \quad (15)$$

In (15), we have defined the normalized amplitudes of the Gaussian mixture $\alpha_i(s) = \beta_i(s)/(\kappa_i |s|^2)$, where the constant κ_i ensures $\sum_{s \in \mathcal{S}_{\mathrm{D}}} \alpha_i(s) = 1$. We also denote the mean and variance of the pdf in (12) by $\mu_{h_j,\mathrm{o}}$ and $\sigma_{h_j,\mathrm{o}}^2$, $j \in \mathcal{P}$, and we define the vector $\boldsymbol{\mu}_{\mathbf{h}}^{\mathrm{o}} = (\mu_{h_i,\mathrm{o}} \mid i \in [1:M+N])^{\mathrm{T}}$ and the matrix $\boldsymbol{\Sigma}_{\mathbf{h}}^{\mathrm{o}}$ with entries $[\boldsymbol{\Sigma}_{\mathbf{h}}^{\mathrm{o}}]_{i,j} = \sigma_{h_i,\mathrm{o}}^2$ if $i = j$ and zero otherwise, for all $i, j \in [1:M+N]$.

Now, using (9) and (14), the message in (13) becomes

$$m_{f_{\mathrm{H}} \to h_i}^{\mathrm{BP}}(h_i) \propto \int \mathrm{CN} \left(\mathbf{h}; \boldsymbol{\mu}_{\mathbf{h}}^{\mathrm{p}}, \boldsymbol{\Sigma}_{\mathbf{h}}^{\mathrm{p}} \right) \mathrm{CN} \left(\mathbf{h}_{\bar{i}}; \boldsymbol{\mu}_{\mathbf{h}_{\bar{i}}}^{\mathrm{o}}, [\boldsymbol{\Sigma}_{\mathbf{h}}^{\mathrm{o}}]_{\bar{i},\bar{i}} \right) \mathrm{d}\mathbf{h}_{\bar{i}}$$
$$\propto \mathrm{CN} \left(h_i; \mu_{h_i,\mathrm{c}}, \sigma_{h_i,\mathrm{c}}^2 \right), \quad (16)$$

with

$$\mu_{h_i,\mathrm{c}} = \mu_{h_i}^{\mathrm{p}} + [\boldsymbol{\Sigma}_{\mathbf{h}}^{\mathrm{p}}]_{i,\bar{i}} \left([\boldsymbol{\Sigma}_{\mathbf{h}}^{\mathrm{o}}]_{\bar{i},\bar{i}} + [\boldsymbol{\Sigma}_{\mathbf{h}}^{\mathrm{p}}]_{\bar{i},\bar{i}} \right)^{-1} (\boldsymbol{\mu}_{\mathbf{h}_{\bar{i}}}^{\mathrm{o}} - \boldsymbol{\mu}_{\mathbf{h}_{\bar{i}}}^{\mathrm{p}}),$$
$$\sigma_{h_i,\mathrm{c}}^2 = [\boldsymbol{\Sigma}_{\mathbf{h}}^{\mathrm{p}}]_{i,i} - [\boldsymbol{\Sigma}_{\mathbf{h}}^{\mathrm{p}}]_{i,\bar{i}} \left([\boldsymbol{\Sigma}_{\mathbf{h}}^{\mathrm{o}}]_{\bar{i},\bar{i}} + [\boldsymbol{\Sigma}_{\mathbf{h}}^{\mathrm{p}}]_{\bar{i},\bar{i}} \right)^{-1} [\boldsymbol{\Sigma}_{\mathbf{h}}^{\mathrm{p}}]_{\bar{i},i}. \quad (17)$$

These messages are further passed as extrinsic values, i.e., $n_{h_i \to f_{\mathrm{D}_i \text{ or } \mathrm{P}_i}}(h_i) = m_{f_{\mathrm{H}} \to h_i}^{\mathrm{BP}}(h_i)$. For each $i \in \mathcal{D}$, the following message is then computed:

$$m_{f_{\mathrm{D}_i} \to x_i}^{\mathrm{BP}}(x_i) \propto \int f_{\mathrm{D}_i}(h_i, x_i) n_{h_i \to f_{\mathrm{D}_i}}(h_i) \, \mathrm{d}h_i$$
$$\propto \frac{1}{\gamma^{-1} + \sigma_{h_i,\mathrm{c}}^2 |x_i|^2} \exp \left(- \frac{|y_i - \mu_{h_i,\mathrm{c}} x_i|}{\gamma^{-1} + \sigma_{h_i,\mathrm{c}}^2 |x_i|^2} \right).$$

After passing the extrinsic messages $n_{x_{i_n} \to f_{\mathrm{M}_n}}(x_{i_n}) = m_{f_{\mathrm{D}_{i_n}} \to x_{i_n}}^{\mathrm{BP}}(x_{i_n})$, $i_n \in \mathcal{D}$, $n \in [1:N]$, we apply the BP update rule to compute the probabilities of the coded and interleaved bits (which is equivalent to MAP demapping), followed by BP decoding to obtain the beliefs of the information bits.

B. Algorithm based on expectation propagation

We set $\mathcal{A}_{\mathrm{MF}} = \emptyset$ and $\mathcal{I}_{\mathrm{EP}} = \{h_i \mid i \in \mathcal{D}\}$. The message $m_{f_{\mathrm{D}_i} \to h_i}^{\mathrm{EP}}(h_i)$ computed with (4) is proportional to a Gaussian pdf; consequently, the EP rule for $m_{f_{\mathrm{H}} \to h_i}^{\mathrm{EP}}(h_i)$ reduces to the BP rule and outputs a Gaussian pdf as in (16), since the operator $\mathcal{G}[\cdot]$ is an identity operator for Gaussian arguments.

Specifically, using (3), (4), and then (11), (16), we have

$$b_{h_i}(h_i) = \mathcal{G} \left[n_{h_i \to f_{\mathrm{D}_i}}(h_i) m_{f_{\mathrm{D}_i} \to h_i}^{\mathrm{BP}}(h_i) \right] = \mathrm{CN} \left(h_i; \mu_{h_i}, \sigma_{h_i}^2 \right),$$

for each $i \in \mathcal{D}$, where

$$\mu_{h_i} = \sum_{s \in \mathcal{S}_{\mathrm{D}}} \phi_i(s) \frac{\sigma_{h_i,\mathrm{c}}^{-2} \mu_{h_i,\mathrm{c}} + \gamma y_i s^*}{\sigma_{h_i,\mathrm{c}}^{-2} + \gamma |s|^2},$$
$$\sigma_{h_i}^2 = \sum_{s \in \mathcal{S}_{\mathrm{D}}} \phi_i(s) \frac{\left| \sigma_{h_i,\mathrm{c}}^{-2} \mu_{h_i,\mathrm{c}} + \gamma y_i s^* \right|^2 + \sigma_{h_i,\mathrm{c}}^{-2} + \gamma |s|^2}{\left(\sigma_{h_i,\mathrm{c}}^{-2} + \gamma |s|^2 \right)^2} - |\mu_{h_i}|^2$$

with

$$\phi_i(s) \triangleq \frac{\beta_i(s)\,\mathrm{CN}\left(y_i;\mu_{h_i,\mathrm{c}}s,\gamma^{-1}+\sigma_{h_i,\mathrm{c}}^{-2}|s|^2\right)}{\sum_{s\in\mathcal{S}_\mathrm{D}}\beta_i(s)\,\mathrm{CN}\left(y_i;\mu_{h_i,\mathrm{c}}s,\gamma^{-1}+\sigma_{h_i,\mathrm{c}}^{-2}|s|^2\right)}$$

and $\mu_{h_i,\mathrm{c}}$, $\sigma_{h_i,\mathrm{c}}^2$ as in (17). Using (4) again, we obtain

$$m_{f_{\mathrm{D}_i}\to h_i}^{\mathrm{EP}}(h_i) \propto \frac{\mathrm{CN}\left(h_i;\mu_{h_i},\sigma_{h_i}^2\right)}{\mathrm{CN}\left(h_i;\mu_{h_i,\mathrm{c}},\sigma_{h_i,\mathrm{c}}^2\right)} \propto \mathrm{CN}\left(h_i;\mu_{h_i,\mathrm{o}},\sigma_{h_i,\mathrm{o}}^2\right),$$

with

$$\begin{aligned} \sigma_{h_i,\mathrm{o}}^{-2} &= \sigma_{h_i}^{-2} - \sigma_{h_i,\mathrm{c}}^{-2}, \\ \mu_{h_i,\mathrm{o}} &= \sigma_{h_i,\mathrm{o}}^2\left(\sigma_{h_i}^{-2}\mu_{h_i} - \sigma_{h_i,\mathrm{c}}^{-2}\mu_{h_i,\mathrm{c}}\right). \end{aligned} \quad (18)$$

Unlike (15) in BP with Gaussian approximation, the values of $\mu_{h_i,\mathrm{o}}$ and $\sigma_{h_i,\mathrm{o}}^2$, $i\in\mathcal{D}$, computed with (18) depend on all $\mu_{h_j,\mathrm{o}}$ and $\sigma_{h_j,\mathrm{o}}^2$, $j\in\mathcal{D}$, $j\neq i$, through (17). The parameters of $m_{f_{\mathrm{H}}\to h_i}^{\mathrm{EP}}(h_i)$ are updated using (17) but with $\mu_{h_i,\mathrm{o}}$ and $\sigma_{h_i,\mathrm{o}}^2$ computed as above. Note that all messages that depend on the channel coefficients need to be updated in a sequential manner. The rest of the messages are computed as in Section IV-A.

C. Algorithm based on the combined BP-MF framework

The factor graph is split into the MF and BP parts by setting $\mathcal{A}_{\mathrm{MF}} = \{f_{\mathrm{D}_i}\mid i\in\mathcal{D}\}$ and $\mathcal{A}_{\mathrm{BP}} = \mathcal{A}\setminus\mathcal{A}_{\mathrm{MF}}$. Such a splitting yields tractable and simple messages, takes advantage of the fact that BP works well with hard constraints and best exploits the correlation between the channel coefficients for the graphical representation in Fig. 1[1].

Assuming we have obtained the messages $n_{x_i\to f_{\mathrm{D}_i}}(x_i)$ (their expression will be given later), we can compute

$$\begin{aligned} m_{f_{\mathrm{D}_i}\to h_i}^{\mathrm{MF}}(h_i) &\propto \exp\left(\int n_{x_i\to f_{\mathrm{D}_i}}(x_i)\ln f_{\mathrm{D}_i}(h_i,x_i)\,\mathrm{d}x_i\right) \\ &\propto \mathrm{CN}\left(h_i;\mu_{h_i,\mathrm{o}},\sigma_{h_i,\mathrm{o}}^2\right), \end{aligned}$$

where

$$\mu_{h_i,\mathrm{o}} = \frac{y_i\mu_{x_i}^*}{\sigma_{x_i}^2+|\mu_{x_i}|^2}, \quad \sigma_{h_i,\mathrm{o}}^2 = \frac{1}{\gamma\left(\sigma_{x_i}^2+|\mu_{x_i}|^2\right)},$$

with the definition $\mu_{x_i}\triangleq\int n_{x_i\to f_{\mathrm{D}_i}}(x_i)\,x_i\,\mathrm{d}x_i$ and $\sigma_{x_i}^2\triangleq\int n_{x_i\to f_{\mathrm{D}_i}}(x_i)|x_i-\mu_{x_i}|^2\,\mathrm{d}x_i$.

The messages $n_{h_i\to f_{\mathrm{H}}}(h_i) = m_{f_{\mathrm{D}_i}\to h_i}^{\mathrm{MF}}(h_i)$ are sent to the BP part and hence are extrinsic values. When computing $m_{f_{\mathrm{H}}\to h_i}^{\mathrm{BP}}(h_i)$ we get the same expression as (16), with the parameters (17). Unlike in the previous algorithms, the following messages are beliefs, i.e., *a posteriori* probabilities (APP):

$$\begin{aligned} n_{h_i\to f_{\mathrm{D}_i}}(h_i) &= \omega_{h_i}\, m_{f_{\mathrm{H}}\to h_i}^{\mathrm{BP}}(h_i)\, m_{f_{\mathrm{D}_i}\to h_i}^{\mathrm{MF}}(h_i) \\ &= \mathrm{CN}\left(h_i;\mu_{h_i},\sigma_{h_i}^2\right), \quad \forall i\in\mathcal{D}, \end{aligned}$$

with

$$\begin{aligned} \mu_{h_i} &= \left(\sigma_{h_i,\mathrm{o}}^{-2}+\sigma_{h_i,\mathrm{c}}^{-2}\right)^{-1}\left(\sigma_{h_i,\mathrm{o}}^{-2}\mu_{h_i,\mathrm{o}}+\sigma_{h_i,\mathrm{c}}^{-2}\mu_{h_i,\mathrm{c}}\right), \quad (19) \\ \sigma_{h_i}^{-2} &= \sigma_{h_i,\mathrm{o}}^{-2}+\sigma_{h_i,\mathrm{c}}^{-2}. \end{aligned}$$

Then, for all $i\in\mathcal{D}$, we compute

$$\begin{aligned} m_{f_{\mathrm{D}_i}\to x_i}^{\mathrm{MF}}(x_i) &\propto \exp\left(\int n_{h_i\to f_{\mathrm{D}_i}}(h_i)\ln f_{\mathrm{D}_i}(h_i,x_i)\,\mathrm{d}h_i\right) \\ &\propto \mathrm{CN}\left(x_i;\frac{y_i\mu_{h_i}^*}{\sigma_{h_i}^2+|\mu_{h_i}|^2},\frac{1}{\gamma(\sigma_{h_i}^2+|\mu_{h_i}|^2)}\right) \quad (20) \end{aligned}$$

and we pass $n_{x_{i_n}\to f_{\mathrm{M}_n}}(x_{i_n}) = m_{f_{\mathrm{D}_{i_n}}\to x_{i_n}}^{\mathrm{MF}}(x_{i_n})$ to the modulation and coding part of the graph as extrinsic values, for all $n\in[1:N]$. After running BP, we obtain (10) and then pass the following APP values back to the MF part:

$$n_{x_{i_n}\to f_{\mathrm{D}_{i_n}}}(x_{i_n}) = \omega_{x_{i_n}}\, m_{f_{\mathrm{M}_n}\to x_{i_n}}^{\mathrm{BP}}(x_{i_n})\, m_{f_{\mathrm{D}_{i_n}}\to x_{i_n}}^{\mathrm{MF}}(x_{i_n}).$$

D. Algorithm based on BP-EM

We now apply EM for channel estimation, so we constrain $b_{h_i}(h_i)$ from the previous BP-MF scheme to be Dirac delta functions. The resulting messages are the same as in the previous subsection, except for $n_{h_i\to f_{\mathrm{D}_i}}(h_i) = \delta(h_i-\mu_{h_i})$ with μ_{h_i} computed as in (19). Note that this algorithm uses only point estimates of the channel weights; however, its complexity is basically still the same, since the computation of (19) actually includes the computation of the corresponding variance.

E. Scheduling of message computations

All algorithms employ the same message-passing scheduling: they start by sending messages $m_{f_{\mathrm{P}_j}\to h_j}(h_j)$ corresponding to pilots and by initializing $m_{f_{\mathrm{D}_i}\to h_i}(h_i)\propto\mathrm{CN}(h_i;0,\infty)$; messages (computed according to the corresponding algorithm) are passed on up to the information bit variables – this completes the first iteration; each following iteration consists in passing messages up to the channel prior factor node and back; messages are passed back and forth until a predefined number of iterations is reached. All algorithms end by taking hard decisions on the beliefs of the information bits.

V. SIMULATION RESULTS

We consider a wireless OFDM system with the parameters given in Table I, and we evaluate by means of Monte Carlo simulations the bit error rate (BER) performance of the receiver algorithms derived in Section IV. We employ as a reference a scheme which has perfect channel state information (CSI), i.e., it has prior knowledge of the vector of channel coefficients $\mathbf{h}$.

We encountered numerical problems with the EP-based scheme due to the instability of EP in general, so we used the heuristic approach [9] to damp the updates of the beliefs b_{h_i} with a step-size $\epsilon = 0.5$. Also, the EP-based scheme has higher computational complexity than the others due to its message definition – it requires multiplication of a Gaussian pdf with a mixture of Gaussian pdfs, the approximation $\mathcal{G}[\cdot]$ and division of Gaussian pdfs – and to the sequentiality of the message updates for the channel coefficients[2].

Results in terms of BER versus signal-to-noise ratio (SNR) are given in Fig. 2, while the convergence of the BER with the number of iterations is illustrated in Fig. 3. The receivers based on EP, combined BP-MF and BP-EM exhibit similar

---

[1]Alternatively, the same level of exploitation of the correlation is obtained by representing the channel variables as a single vector variable $\mathbf{h}$ and "moving" factor node f_{H} to the MF part [11].

[2]For the other receiver schemes, it can be shown that the parameters of all messages $m_{f_{\mathrm{H}}\to h_i}^{\mathrm{BP}}(h_i)$ with $i\in\mathcal{D}\cup\mathcal{P}$ can be computed jointly and with a lower complexity.
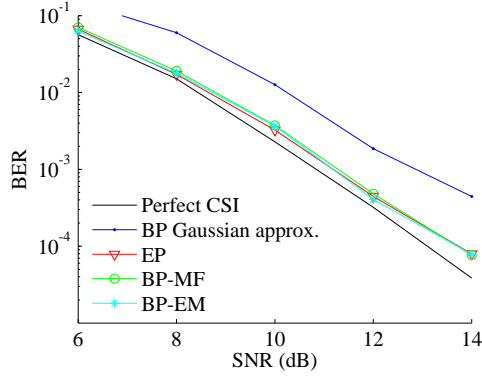
Fig. 2. BER vs. SNR performance of the receiver algorithms for a number of pilot symbols $M = 10$, corresponding to a high pilot spacing $\Delta_\mathrm{P} \approx 2.5 W_\mathrm{coh}$.
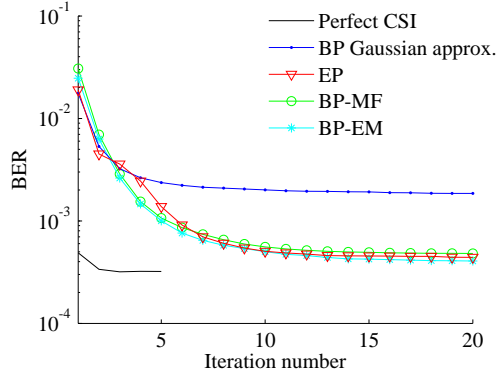

Fig. 3. Convergence of the BER performance from Fig. 2 at SNR $= 12$ dB.

performance. They significantly outperform the receiver employing BP with Gaussian approximation. Note that even with a high pilot spacing $\Delta_\mathrm{P} \approx 2.5 W_\mathrm{coh}$ the performance of the former algorithms is close to that of the receiver having perfect CSI. These three algorithms converge in about 10–12 iterations, while BP with Gaussian approximation converges a little faster, but to a higher BER value. Other results not presented here show that for a higher pilot density the algorithms converge faster, as expected.

Note that the results for the (essentially equally-complex) BP-EM and BP-MF receivers are nearly identical, even if the former discards the soft information in channel estimation. We noticed during our evaluations that $\sigma_{h_i}^2 \ll |\mu_{h_i}|^2$ even at low SNR values, so our explanation would be that accounting for $\sigma_{h_i}^2$ in the BP-MF receiver does not have a noticeable impact on the detection (20).

TABLE I
PARAMETERS OF THE WIRELESS OFDM SYSTEM

Parameter	Value
Subcarrier spacing	$15 \, \mathrm{kHz}$
Number of active subcarriers	$M + N = 300$
Number of evenly spaced pilot symbols	$M = 10$
Pilot spacing	$\Delta_\mathrm{P} \approx 500 \, \mathrm{kHz}$
Modulation scheme for data symbols	$16\,\mathrm{QAM}$ ($L = 4$)
Convolutional channel code	$R = 1/3$, $(133, 171, 165)_8$
Multipath channel model	3GPP ETU
Coherence bandwidth of the channel	$W_\mathrm{coh} \approx 200 \, \mathrm{kHz}$

VI. CONCLUSIONS

We formulated the problem of joint channel estimation and decoding in a communication system as inference in a graphical model. To solve the inference problem, we resorted to a recently proposed message-passing framework that unifies the BP and MF algorithms and includes them as particular instances. Additionally, we illustrated how the combined framework can encompass the EP and EM inference algorithms.

Based on the inference techniques considered, we derived four receiver algorithms. Since BP is not suitable for the studied problem, as it leads to intractable messages, we applied its variant which employs Gaussian approximation of the computationally cumbersome messages instead. However, our results showed that it performs significantly worse than the other proposed schemes. Considering the BER results, the computational complexity and stability of these schemes, we conclude that the receiver based on the combined BP-MF framework and its BP-EM variant are the most effective receiver algorithms.

ACKNOWLEDGMENT

REFERENCES

[1] M. Tüchler and A. C. Singer, "Turbo equalization: An overview," *IEEE Transactions on Information Theory*, vol. 57, no. 2, pp. 920–952, 2011.

[2] J. Boutros and G. Caire, "Iterative multiuser joint decoding: unified framework and asymptotic analysis," *Information Theory, IEEE Transactions on*, vol. 48, no. 7, pp. 1772–1793, jul 2002.

[3] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Foundations and Trends in Machine Learning*, vol. 1, pp. 1–305, 2008.

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

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

[6] J. Winn and C. Bishop, "Variational message passing," *Journal of Machine Learning Research*, vol. 6, pp. 661–694, 2005.

[7] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[8] T. Minka, "Expectation propagation for approximate bayesian inference," in *Proc. 17th Conf. on Uncertainty in AI*, 2001, pp. 362–369.

[9] ——, "Divergence measures and message passing," Microsoft Research, Tech. Rep., 2005.

[10] J. Yedidia, W. Freeman, and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms," *IEEE Trans. Inform. Theory*, vol. 51, no. 7, pp. 2282–2312, July 2005.

[11] E. Riegler, G. E. Kirkelund, C. N. Manchón, M.-A. Badiu, and B. H. Fleury, "Merging belief propagation and the mean field approximation: A free energy approach," *accepted for publication in IEEE Trans. Inform. Theory, 2012, arXiv:1112.0467v2[cs.IT]*.

[12] Z. Shi, T. Wo, P. Hoeher, and G. Auer, "Graph-based soft iterative receiver for higher-order modulation," in *Proc. 12th IEEE Int. Conf. on Comm. Tech. (ICCT)*, Nanjing, China, Nov. 2010, pp. 825–828.