An Iterative Receiver for OFDM With Sparsity-Based Parametric Channel Estimation

Hansen, Thomas Lundgaard; Jørgensen, Peter Bjørn; Badiu, Mihai Alin; Fleury, Bernard Henri

Published in:
I E E E Transactions on Signal Processing

DOI (link to publication from Publisher):
10.1109/TSP.2018.2868314

Publication date:
2018

Document Version
Accepted author manuscript, peer reviewed version

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2868314, IEEE Transactions on Signal Processing

An Iterative Receiver for OFDM With Sparsity-Based Parametric Channel Estimation

Thomas L. Hansen, Peter B. Jørgensen, Mihai-Alin Badiu and Bernard H. Fleury

***Abstract*—In this work we design a receiver that iteratively passes soft information between the channel estimation and data decoding stages. The receiver incorporates sparsity-based parametric channel estimation. State-of-the-art sparsity-based iterative receivers simplify the channel estimation problem by restricting the multipath delays to a grid. Our receiver does not impose such a restriction. As a result it does not suffer from the leakage effect, which destroys sparsity. Communication at near capacity rates in high SNR requires a large modulation order. Due to the close proximity of modulation symbols in such systems, the grid-based approximation is of insufficient accuracy. We show numerically that a state-of-the-art iterative receiver with grid-based sparse channel estimation exhibits a bit-error-rate floor in the high SNR regime. On the contrary, our receiver performs very close to the perfect channel state information bound for all SNR values. We also demonstrate both theoretically and numerically that parametric channel estimation works well in dense channels, i.e., when the number of multipath components is large and each individual component cannot be resolved.**

***Index Terms*—Iterative receivers, message-passing algorithms, sparse channel estimation, parametric channel estimation, off-the-grid compressed sensing.**

I. Introduction

Achieving high data-rate wireless communication with large spectral efficiency requires the use of higher-order modulation formats, e.g. up to 256-QAM in 3GPP LTE [1]. Clearly using a high modulation order presuppose a large signal-to-noise ratio (SNR), which will be supported by the envisioned transition to small-cell operation. The availability of channel estimation schemes that achieve high accuracy is crucial for receivers of systems with large modulation order operating in the high-SNR regime.

To facilitate channel estimation, current systems embed pilot symbols into the transmitted signal. In orthogonal frequency-division multiplexing (OFDM) systems, a number of sub-carriers are assigned to transmit pilot symbols. The number of pilots is chosen to optimize throughput as a trade-off between the amount of bandwidth and power allocated to pilot transmission and fidelity of the channel estimate.

In this work we seek to improve upon this trade-off by designing a highly accurate channel estimator while requiring a low pilot overhead. We propose a unified receiver design that incorporates two main ideas: *a)* an iterative architecture and *b)* sparsity-based parametric channel estimation.

T. L. Hansen, M.-A. Badiu and B. H. Fleury are with the Department of Electronic Systems at Aalborg University, Denmark. P. B. Jørgensen is with the Technical University of Denmark; this work was conducted while he was with Aalborg University. Contact e-mail: tlh@es.aau.dk.

This work was supported by the Danish Council for Independent Research under grant ids DFF–4005-00549 and DFF–5054-00212 and by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMmunications NEWCOM# (Grant agreement no. 318306). This work was also supported by the cooperative research project VIRTUOSO, funded by Intel Mobile Communications, Anite, Telenor, Aalborg University and Innovation Fund Denmark.

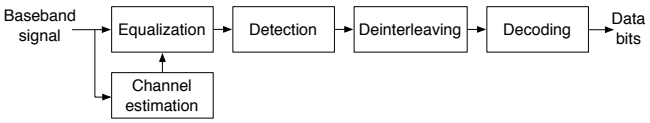

Fig. 1. Flowchart of classical receiver design.

Our proposed receiver does not require any a-priori statistical information about the wireless channel and is thus a particularly good candidate for systems where no such information is available.

A. Design of Iterative Receivers

Classical receiver design employs a functional splitting of the process in the receiver into independent subtasks, as illustrated in Fig. 1. Such a structure is suboptimal, since the information learned from the received signal in any of the subtasks is only utilized in subsequent subtasks. To remedy this sub-optimality feedback loops can be introduced between the functional blocks in the receiver. This approach is known as the turbo principle [2], [3] due to its resemblance to iterative decoding of turbo codes.

Application of the turbo principle has led to many iterative receiver designs, e.g. [2]–[4]. Common to these works is that each of the subtasks are designed independently using traditional methods such as maximum likelihood (ML), maximum a-posteriori probability (MAP) or minimum mean squared error (MMSE). The work [5] introduced receiver design from the perspective of inference in a factor graph. This allows for the receiver subtasks to be designed *jointly* with a certain objective in mind; a common example is to seek the MAP estimate of the information bits. Due to tractability and computational constraints, approximate inference methods must be employed for iterative receiver design. Examples of such methods are expectation propagation [6], belief propagation (BP) with approximated messages [7], combined BP and mean-field (MF) [8], [9], relaxed BP [10] and generalized approximate message-passing (GAMP) [11].

B. Parametric Channel Estimation

The impulse response of the compound channel (composed of the transmitter RF front-end, the propagation channel and the receiver RF front-end) is traditionally modelled as a sum of the form

$$g(\tau) = \sum_{l=1}^{L} \alpha_l c(\tau - \tau_l), \tag{1}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2868314, IEEE Transactions on Signal Processing

where $c(\tau)$ denotes the compound impulse response of the transmitter and receiver RF front-ends. Here, L is the number of so-called multipath components. The lth multipath component is characterized by its coefficient or weight $\alpha_l \in \mathbb{C}$ and its relative delay $\tau_l \in \mathbb{R}$. For short we refer to $g(\tau)$ in (1) as the channel impulse response (CIR). When the number of multipath components L is small relative to the number of OFDM subcarriers, the model (1) provides a parsimonious representation of the compound channel and it is advantageous to perform channel estimation by estimating the parameters of this model, i.e., estimating L, α_l and τ_l for $l = 1, \ldots, L$. We refer to this approach as parametric channel estimation. It has been generally understood for many years [12], [13] that the delays can only be estimated when the multipath components are well separated (see footnote 7).

In our application context it is very restrictive to assume that the CIR takes the form (1) with L small and all delays well separated. In Sec. II-B we demonstrate that even if the wireless channel exhibits a very large number of (closely located) multipath components its CIR can still be approximated by (1) with L small[1]. We refer to this approximation as the *virtual* CIR. We show that the corresponding virtual channel frequency response (CFR) accurately approximates the actual CFR within the system bandwidth. This means that we can use (1) with L small as an estimation model even for channels of which the multipath components cannot be resolved with the used system bandwidth.

Early works on parametric channel estimation address applications to underwater communications [14] and ultra-wideband (UWB) communications [15], [16]. Another classical example is the rake receiver [17]. All of these older works assume that the number of (virtual) multipath components L is known a priori or use heuristics to estimate it.

A sparsity-based (or compressed sensing-based) approach can be used to allow for inherent estimation of the number of (virtual) multipath components. Most literature on sparsity-based channel estimation [18]–[24] employs a grid-based approximation of the CIR model (1), where the multipath delays are confined to a discrete set of possible values. When a baud-spaced grid[2] is used, we refer to the samples of the CIR (1) as channel taps. The grid-based approximation results in a leakage effect[3] [21], [25] and the vector of channel taps is therefore only approximately sparse [11], [23], [24]. We demonstrate in our numerical investigation that the grid-based approximation impairs the performance of receivers for OFDM systems with large modulation order operating in the high-SNR regime. From a compressed sensing point of view the effect of the grid-based approximation can be understood as a basis mismatch [26].

Recent works on off-grid compressed sensing have proposed methods that could in principle be applied to sparsity-based channel estimation without resorting to the grid approximation. These are based on atomic-norm minimization [13], [27], [28], finite rate of innovation [29] or Bayesian inference [30]–[33]. While all these methods show good performance, the former two cannot easily be incorporated in an iterative receiver. In this paper we show how sparsity-based parametric channel estimation can be incorporated in an iterative receiver by using approximate Bayesian inference. Our channel estimation scheme is *sparsity-based* in the sense that a sparsity-promoting prior model is used to achieve inherent estimation of the number of (virtual) multipath components (the vector $\boldsymbol{z}$ associated with (12) is sparse) and it is *parametric* in the sense that a parametric channel model is used to design the channel estimator.

C. Prior Art

Several prior works incorporate sparsity-based channel estimation in an iterative receiver. Prasad et. al. [23], [24] propose a joint sparse channel estimation and *detection* scheme for OFDM transmission. Channel decoding is not considered in the joint processing and the EM algorithm is used for channel inference. A baud-spaced grid is used.

Iterative receiver design for OFDM systems via GAMP and relaxed BP is proposed by Schniter in [10], [11]. The estimated multipath delays are restricted to the baud-spaced grid. In the numerical evaluation of [10] the CIRs fulfill this restriction, thus avoiding the leakage effect at the expense of introducing an unrealistic channel model. In [11] a channel model generating continuous-valued delays is assumed. It is shown that the channel taps follow a super-Gaussian density that is modelled via a two-component Gaussian mixture. Due to the baud-spaced grid the channel taps are correlated, which is mimicked with a hidden Markov model. The resulting model has a large number of parameters to be estimated, that causes systems with high-order modulation format to exhibit a bit-error-rate (BER) floor when operating in the high-SNR regime (see Sec V).

The problem of parametric channel estimation based only on pilots or in the contrived case when the data symbols are given is equivalent to that of line spectral estimation [22]. The work [31] proposes a variational Bayesian approach to line spectral estimation. It is shown that the Bernoulli-Gaussian prior [34] is a powerful and tractable sparsity-inducing model. Our sparsity-based parametric channel estimator is inspired by [31] and uses the Bernoulli-Gaussian prior model too. It differs from [31] in several aspects: *a)* at the data subcarriers the observations are modulated with the unknown data symbols, *b)* we impose that the estimate of the posterior probability density function (pdf) of the multipath coefficients factorizes and *c)* to reduce computational complexity we use a point estimate of the multipath delays.[4]

D. Contributions

The contributions of this paper are as follows:

[1] Our analysis makes the usual assumption in OFDM of time-limited CIR, see (3). Our results are therefore only directly applicable to scenarios where this assumption can reasonably be made, as is usually the case in radio communication.

[2] In the baud-spaced grid, the distance between adjacent grid points is the reciprocal of the system bandwidth.

[3] The compound wireless channel has a representation on the baud-spaced grid obtained by sampling $g(\tau)$. This representation is not sparse due to the presence of the RF front-end filter $c(\tau)$ in (1) that introduces leakage.

[4] By contrast, the scheme in [31] applied in our context computes estimates of the posterior distribution of the delays.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2868314, IEEE Transactions on Signal Processing

1) We propose a method to incorporate sparsity-based parametric channel estimation into an iterative receiver. Specifically we use the combined BP and MF (BP-MF) framework [8] to derive such an iterative receiver within a unified framework.
2) We show (numerically) that iterative receivers for OFDM with high modulation order exhibit an error floor in the high-SNR regime when they employ state-of-the-art sparse channel estimation based on the baud-spaced grid approximation. Our iterative receiver design demonstrates how this error floor can be avoided.
3) We demonstrate that parametric channel estimation, contrary to immediate intuition, can be applied to both specular and dense channels. In particular it is shown that the impulse response of any uncorrelated scattering channel can be approximated within the system bandwidth by a virtual CIR of the form (1). The number of components L in the virtual CIR is equal to the effective rank of the channel covariance matrix. We demonstrate numerically that the effective rank of the channel covariance matrix is low for both a specular and a dense synthetic channel.
4) Our algorithm development demonstrates how the BP-MF framework can be modified to provide approximate ML estimation of model parameters and how some latent variables can be estimated jointly to improve convergence speed. We expect that these approaches will prove useful in other applications of BP-MF.

Our receiver only uses a few parameters (specifically the noise variance and the two parameters of the Bernoulli-Gaussian prior model, sparsity level ρ and multipath coefficient variance η) to describe the statistical properties of the CIR and these are inherently estimated by appropriately modifying BP-MF. This is in contrast to, for example, the linear MMSE (LMMSE) channel estimators, which require a-priori specification of the second-order statistics of the CFR [4], [35], and the GAMP receiver [10], which relies on the second-order statistics of the channel taps and the transition probabilities of the hidden Markov model.

The parametric channel estimation scheme that we propose requires the compound frequency response of the RF front-ends to be known (at active subcarriers). While that is not a wholly unrealistic assumption, it may prove too restrictive in some practical situations. Since this frequency response is stable over many OFDM symbols we expect that it can be estimated; however that falls outside the scope of this work.

E. Notation and Outline

We denote column vectors as $\boldsymbol{a}$ and matrices as $\mathbf{A}$. Transposition is denoted as $(\cdot)^{\mathrm{T}}$ and conjugate (Hermitian) transposition as $(\cdot)^{\mathrm{H}}$. The scalar a_i or $[\boldsymbol{a}]_i$ gives the ith entry of vector $\boldsymbol{a}$, while $\boldsymbol{a}_{\mathcal{S}}$ gives a vector containing the entries in $\boldsymbol{a}$ at the indices in the integer set $\mathcal{S}$. The set difference operator $\mathcal{S}\backslash\{i\}$ gives the index set $\mathcal{S}$ with index i removed; we abuse notation slightly and write $\mathcal{S}\backslash i$ for short. The notation $[\mathbf{A}]_{i,k}$ gives the (i,k)th element of matrix $\mathbf{A}$. We denote the vector $\boldsymbol{a}$ with the ith element removed as $\boldsymbol{a}_{\backslash i}$ and use a similar notation for matrices with columns and/or rows removed (e.g. $[\mathbf{A}]_{i,\backslash k}$ for the ith row with kth entry removed). The notation $\mathrm{diag}(\boldsymbol{a})$ denotes a matrix with the entries of $\boldsymbol{a}$ on the diagonal and zeros elsewhere. The indicator function $\mathbb{1}_{[\cdot]}$ gives 1 when the condition in the brackets is fulfilled and 0 otherwise. The notation $a \propto^{\mathrm{e}} b$ denotes $\exp(a) \propto \exp(b)$, which implies $a = b + \mathrm{const}$. The multivariate complex normal probability density function (pdf) is defined as

$$\mathrm{CN}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) \triangleq \pi^{-\dim(\boldsymbol{x})}|\boldsymbol{\Sigma}|^{-1}\exp\left(-(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{H}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right).$$

The notation $\mathrm{unif}(x;0,T)$ gives the continuous uniform pdf on the interval $[0,T]$ and $\mathrm{Bern}(x;\rho)$ gives the Bernoulli probability mass function (pmf) for $x \in \{0,1\}$ with probability of success ρ. We use $*$ to denote convolution and $\delta(\cdot)$ and $\delta[\cdot]$ to denote the Dirac and Kronecker delta, respectively.

The paper is structured as follows: In Section II we specify the observation model. In Section III our approach to approximate Bayesian inference is discussed. The inference algorithm is derived in detail in Section IV. Section V presents the numerical evaluation. Conclusions are given in Section VI.

II. Modelling

We consider data transmission using a single-input single-output OFDM system. Since we do not exploit any structure between consecutive OFDM symbols, we model the sequence of transmitted OFDM symbols to be independent and identically distributed (i.i.d.). The OFDM system transmits P pilot subcarriers and D data subcarriers, such that the total number of subcarriers per symbol is $N = P + D$. The sets $\mathcal{P}$ and $\mathcal{D}$ give the indices of the pilot and data subcarriers, respectively. It follows that $\mathcal{D} \cup \mathcal{P} = \{1,\ldots,N\}$ and $\mathcal{D} \cap \mathcal{P} = \emptyset$.

A. OFDM System

The K (equi-probable) information bits to be transmitted are stacked in vector $\boldsymbol{u} \in \{0,1\}^K$. These bits are coded by a rate-R encoder and interleaved to get the length-K/R vector $\boldsymbol{c} = \mathcal{C}(\boldsymbol{u})$. The interleaving and coding function $\mathcal{C} : \{0,1\}^K \to \{0,1\}^{K/R}$ can represent any interleaver and coder, e.g. a turbo [36], low-density parity check (LDPC) [37] or convolutional code. We split $\boldsymbol{c}$ into subvectors $\boldsymbol{c}^{(i)} \in \{0,1\}^Q$, $i \in \mathcal{D}$, such that $\boldsymbol{c}^{(i)}$ contains the Q bits that are mapped to the ith subcarrier. The complex symbols $x_i = \mathcal{M}(\boldsymbol{c}^{(i)})$, $i \in \mathcal{D}$, are obtained via the 2^Q-ary mapping $\mathcal{M} : \{0,1\}^Q \to \mathbb{A}_{\mathrm{D}} \subset \mathbb{C}$, where $\mathbb{A}_{\mathrm{D}}$ is the data symbol alphabet. The pilots are selected in the pilot symbol alphabet $\mathbb{A}_{\mathrm{P}} \subset \mathbb{C}$. In OFDM, $\mathbb{A}_{\mathrm{D}}$ is typically a 2^Q-ary quadrature amplitude modulation (QAM) alphabet and $\mathbb{A}_{\mathrm{P}}$ a quadrature phase shift keying (QPSK) alphabet. The pilot and data symbols are stacked in vector $\boldsymbol{x}$. Vector $\boldsymbol{x}_{\mathcal{D}}$ contains the data symbols and $\boldsymbol{x}_{\mathcal{P}}$ contains the pilot symbols.

The transmitter and receiver are assumed to operate with perfect time synchronization. We also assume that the local oscillators in the transmitter and receiver are perfectly synchronized and that these oscillators are ideal (i.e., no phase noise, etc.) We consider a baseband signal model and assume ideal conversion to and from the carrier-frequency passband signal. The RF front-ends are modelled as linear time-invariant filters with compound impulse response $c(\tau) = c_{\mathrm{TX}}(\tau) * c_{\mathrm{RX}}(\tau)$. The wireless channel is also assumed linear and time-invariant for

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2868314, IEEE Transactions on Signal Processing

the duration of an OFDM symbol. The impulse response of the propagation channel during transmission of the current OFDM symbol is denoted $h(\tau)$ and the (compound) CIR is then

$$g(\tau) = c(\tau) * h(\tau). \tag{2}$$

We make the usual assumption of time-limited CIR:

$$g(\tau) = 0 \quad \text{for} \quad \tau \notin [0, T_{\text{CP}}], \tag{3}$$

where T_{CP} is the cyclic prefix duration. In practice this assumption needs only to be fulfilled relative to the noise level.[5]

By the assumption in (3) the OFDM system operates without inter-symbol interference, so we can consider transmission of a single OFDM symbol. The OFDM transmitter is modelled as a baseband processor followed by an RF front-end that applies the filter $c_{\text{TX}}(\tau)$. The baseband processor emits

$$s(t) = \begin{cases} \sum_{n=1}^{N} x_n \exp(j2\pi\Delta_f n t) & t \in [-T_{\text{CP}}, T_{\text{sym}}] \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

where Δ_f gives the subcarrier spacing and $T_{\text{sym}} = \Delta_f^{-1}$ is the OFDM symbol length. The OFDM receiver is modelled as an RF front-end that applies the filter $c_{\text{RX}}(\tau)$ followed by a baseband processor that samples the signal. The signal at the output of the receiver RF front-end is

$$r(t) = g(\tau) * s(t) + w(t), \tag{5}$$

where $w(t)$ is low-pass filtered white Gaussian noise. The receiver baseband processor samples $r(t)$, removes the cyclic prefix and calculates the discrete Fourier transform to obtain the observed vector $\boldsymbol{y}$. The assumption in (3) ensures that orthogonality of the subcarriers is preserved. It can be shown [38] that

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{g} + \boldsymbol{w}, \tag{6}$$

where $\mathbf{X} = \text{diag}(\boldsymbol{x})$. The Gaussian noise vector $\boldsymbol{w}$ is assumed[6] white with component variance β. The vector $\boldsymbol{g}$ contains samples of the compound CFR at the subcarrier frequencies and its entries are

$$g_n = \int_0^{T_{\text{sym}}} g(\tau) \exp(-j2\pi\Delta_f n\tau)\, \mathrm{d}\tau, \quad n = 1, \ldots, N. \tag{7}$$

Inserting (2) into (7) and by the convolution theorem we can obtain (see [38] for details)

$$\boldsymbol{g} = \mathbf{C}\boldsymbol{h}, \tag{8}$$

where $\mathbf{C} = \text{diag}(\boldsymbol{c})$. The vectors $\boldsymbol{c}$ and $\boldsymbol{h}$ contain samples of the Fourier transform of $c(\tau)$ and $h(\tau)$, respectively. These vectors are obtained analogously to (7).

B. Parametric Channel Model

We now consider a model for the propagation channel $h(\tau)$. A classical model is the uncorrelated-scattering (US) channel [39], in which $h(\tau)$ is modelled as a stochastic process with autocorrelation

$$\mathbb{E}[h(\tau)h^*(\tau')] = \rho(\tau)\delta(\tau - \tau'). \tag{9}$$

The function $\rho(\tau)$ is the power-delay profile (PDP). We further assume that the process $h(\tau)$ is zero-mean. The vector $\boldsymbol{h}$ is then also zero-mean. Denote the covariance matrix of $\boldsymbol{h}$ as $\boldsymbol{\Sigma} = \mathbb{E}\big[\boldsymbol{h}\boldsymbol{h}^{\text{H}}\big]$. Using the US assumption it can be shown that the frequency-response vector $\boldsymbol{h}$ contains samples of a wide-sense-stationary random process and that $\boldsymbol{\Sigma}$ is a Toeplitz matrix.

Denote the rank of the $N \times N$ matrix $\boldsymbol{\Sigma}$ as L. Then the Caratheódory parameterization of a Toeplitz matrix [40], [41] states that there exist vectors $\boldsymbol{\tau} \in [0, \Delta_f^{-1})^L$ and $\boldsymbol{\gamma} \in [0, \infty)^L$ such that

$$\boldsymbol{\Sigma} = \boldsymbol{\Psi}(\boldsymbol{\tau})\boldsymbol{\Gamma}\boldsymbol{\Psi}(\boldsymbol{\tau}), \tag{10}$$

where $\boldsymbol{\Gamma} = \text{diag}(\boldsymbol{\gamma})$ and the matrix $\boldsymbol{\Psi}(\boldsymbol{\tau}) \in \mathbb{C}^{N\times L}$ has (n,l)th entry $\exp(-j2\pi\Delta_f n\tau_l)$, $n = 1, \ldots, N$, $l = 1, \ldots, L$. Note that the parameterization is unique if and only if $L < N$. From (10) it is clear that $\boldsymbol{h}$ lies in the column space of $\boldsymbol{\Psi}(\boldsymbol{\tau})$ and that it can be represented as $\boldsymbol{h} = \boldsymbol{\Psi}(\boldsymbol{\tau})\boldsymbol{\alpha}$ for some $\boldsymbol{\alpha} \in \mathbb{C}^L$. It then follows that

$$\boldsymbol{g} = \mathbf{C}\boldsymbol{\Psi}(\boldsymbol{\tau})\boldsymbol{\alpha}. \tag{11}$$

The parametric channel estimator that we employ is obtained by estimating $\boldsymbol{\tau}$ and $\boldsymbol{\alpha}$ in the above parametric model of $\boldsymbol{g}$. It is recognized that $\boldsymbol{h}$ is a superposition of complex sinusoids. Thus, given $\mathbf{X}$ and $\mathbf{C}$, the estimation of L, $\boldsymbol{\alpha}$ and $\boldsymbol{\tau}$ reduces to an instance of line spectral estimation.

The reuse of notation between (1) and (11) is not accidental. If the CIR is assumed to take the parametric from (1) and this CIR is Fourier transformed to obtain the CFR, we get exactly the expression (11). Parametric channel estimators are in fact usually motivated by assuming that the CIR has the from (1). But in the above we showed that the parametric model (11) can be obtained from the US assumption, i.e., without explicitly imposing a model of the form (1). This means that parametric channel estimation can be used for all US channels. The pair (τ_l, α_l) denotes the delay and complex coefficient of a virtual multipath component. The L virtual multipath components described by $(\boldsymbol{\tau}, \boldsymbol{\alpha})$ can be inserted into (1) to obtain a virtual CIR. The above shows that if the covariance matrix $\boldsymbol{\Sigma}$ indeed has rank L the corresponding virtual CFR coincides with the CFR of the actual channel within the system bandwidth.

In this work we make the simplifying assumption that the filters in the RF front-ends have constant frequency response within the system bandwidth. This assumption means that $\mathbf{C} = \mathbf{I}$ (any constant scaling can be integrated into $\boldsymbol{\alpha}$). This assumption is reasonable because typical OFDM systems employ a number of unused virtual (or guard) subcarriers in the roll-off region of the RF front-end filters [42].

Remark 1: The assumption $\mathbf{C} = \mathbf{I}$ does not mean that there are no filters at RF front-ends. It just means that these filters have unit frequency response within the system bandwidth. To be precise, the wireless channel is "observed" by the receiver as described by (6) and (11). It is clear that the wireless

[5]Specifically the signal contribution in (6) arising from the tail of the compound CIR outside $[0, T_{\text{CP}}]$ should be neglectable compared to noise.

[6]This assumption is fulfilled when the receive RF front-end has constant frequency response within the system bandwidth.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2868314, IEEE Transactions on Signal Processing

channel is observed only within a band-limited interval of length $N\Delta_f$.

Remark 2: The assumption $\mathbf{C} = \mathbf{I}$ can be relaxed to the assumption that the frequency response of the RF front-ends is arbitrary but known or estimated by the receiver (i.e., the matrix $\mathbf{C}$ is known or estimated) [43]. Since $\mathbf{C}$ is fixed across many OFDM symbols we expect that it can be estimated with high accuracy by the receiver. An investigation of such an approach is outside the scope of this paper. If $\mathbf{C}$ is known the derivation in Sec. IV can be straightforwardly extended to include $\mathbf{C}$. We use $\mathbf{C} = \mathbf{I}$ to keep the notation simple.

C. Specular and Dense Channels

The rank of the channel covariance matrix L describes the channel's number of degrees of freedom. The smaller this number, the fewer parameters are needed in (11) to describe the channel and the higher channel estimation accuracy can be achieved. We are thus particularly interested in the case where $\boldsymbol{\Sigma}$ is low-rank.

In this paper we classify channels into the two categories of specular and dense channels. For specular channels the CIR truly has the form (1) with L much smaller than N and the delays in $\boldsymbol{\tau}$ are well separated.[7] In such channels the delays $\boldsymbol{\tau}$ and coefficients $\boldsymbol{\alpha}$ can directly be estimated as indicated by [13]. It is easy to show that the channel covariance matrix of a specular channel does indeed take the form (10) and that it has low rank. Empirical evidence suggests that the wireless channel in some propagation environments is specular to a large extent. In practice, specular channels are composed of a small number of dominant multipath components and a remaining part with power below the noise floor. Examples include the ultra-wideband channels that are considered for 5G wireless communications [44], [45] and underwater acoustic channels [46]. See also [18], [47] and references therein.

It is, however, broadly accepted that wireless channels are not always specular [11], [18], [48]. In the general case they are composed of a very large number of multipath components that do not adhere to a minimum separation condition. That is caused by diffuse scattering and by rich scattering environments. We refer to such channels as dense. In dense channels it is not possible to estimate the delay and coefficient of each multipath component in (1). The use of the Caratheódory parameterization shows that it is, however, still possible to estimate a set of virtual multipath components that approximate the actual CFR within the system bandwidth. As discussed above the parametric approach works better when $\boldsymbol{\Sigma}$ has low-rank or, in other words, when the virtual CIR has only few components. Using a representation based on discrete prolate spheroidal sequences [49] it can be shown that the assumption (3) implies that $\boldsymbol{h}$ effectively lies in a subspace with dimension approximately given by $\lceil T_{\text{CP}} N \Delta_f \rceil$. This value then also gives an upper bound on the effective rank[8] of $\boldsymbol{\Sigma}$. OFDM systems are practically always designed such that $T_{\text{CP}}\Delta_f \ll 1$ and so $\boldsymbol{\Sigma}$ has low effective rank. In many cases the effective rank is even lower than $\lceil T_{\text{CP}} N \Delta_f \rceil$. We demonstrate below that this is the case for a standardized and widely used model of a dense channel.

[7] "Well separated" is here meant relative to the reciprocal of the system bandwidth $1/(N\Delta_f)$.

[8] By the effective rank we mean a rank that ignores very small eigenvalues.

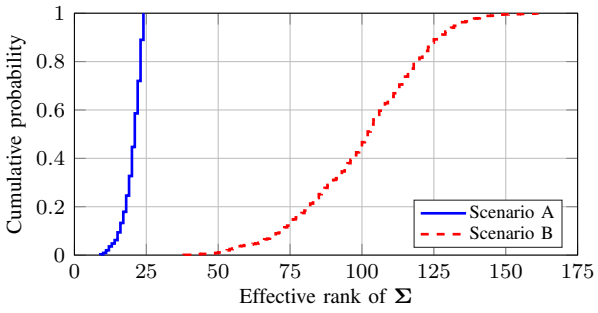

Fig. 2. Empirical CDFs of effective rank of $\boldsymbol{\Sigma}$ when a threshold is applied to the eigenvalues. The plot is obtained by estimating the CDF from $1,000$ realizations in each of the scenarios described in Sec. V.

To investigate the effective rank of the channel covariance matrix we conduct a numerical experiment. We here give a quick overview of this experiment and refer to [50, Chap. 3] for more details. The experiment is based on the two propagation scenarios described in Sec. V. The channel model of Scenario A is specular to a large extent while the channel model of Scenario B is dense. The channel models randomly generate a set of multipath delays and component powers that describe the local behaviour of the channel. The channel covariance matrix is obtained by inserting these delays and powers into (10). The obtained $\boldsymbol{\Sigma}$ has many eigenvalues that are small but still non-zero (this is not caused by limited numerical precision). We therefore define a method to neglect small eigenvalues that do not represent significant power. For that purpose the eigenvalues are normalized such that their average value is 1. The normalized eigenvalues below 10^{-4} are then set to zero. This means that whenever the SNR is significantly below $40\,\text{dB}$, the removed power is neglectable in comparison to the noise power. Even in perfect conditions wireless communication systems practically always operate significantly below 40 dB SNR. For that reason we find this to be a conservative approach to thresholding the eigenvalues. Taking the resulting number of non-zero eigenvalues gives the effective rank. Fig. 2 depicts empirical cumulative distribution functions (CDFs) of the effective rank in Scenario A and B. From Table I we have $\lceil T_{\text{CP}} N \Delta_f \rceil = 134$ in Scenario A and $\lceil T_{\text{CP}} N \Delta_f \rceil = 205$. It is seen that the effective rank of $\boldsymbol{\Sigma}$ is generally much smaller than $\lceil T_{\text{CP}} N \Delta_f \rceil$. In Scenario B the effective rank is also much smaller than the number of multipath components in the channel, indicating that a virtual CIR of the form (1) with L small exists even for dense channels. Due to the use of the effective rank (and not the true rank) of the channel covariance matrix, the virtual CFR approximates the true CFR. The approximation is only accurate within the system bandwidth.

In summary it can be concluded that parametric channel estimation can be applied to both specular and dense channels. That is indeed confirmed in the numerical investigation reported in Sec. V.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2868314, IEEE Transactions on Signal Processing

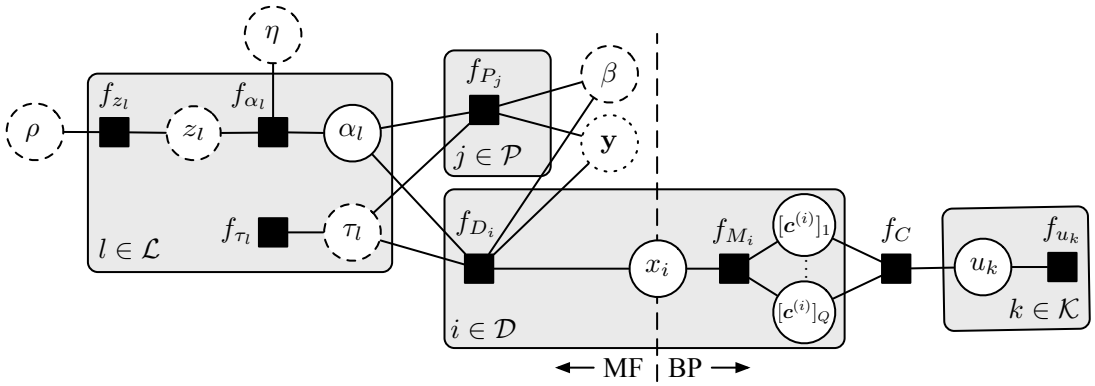

Fig. 3. Factor graph representation of the probabilistic model describing the complete OFDM system and channel model. The shaded areas indicate multiple copies of the nodes, as specified by the index sets. The vector of observations $\boldsymbol{y}$ is included with a dotted line because it is known at the time of inference. Variables of which a point estimate is obtained (as opposed to a full variational estimate of the posterior pdf) are represented by circles with dashed line. The vertical dashed line shows the separation between the BP and MF subgraphs.

D. Probabilistic Model of the OFDM System

We are now ready to present a probabilistic model that describes the complete OFDM system. The model expresses the joint probability of all variables in the system as a product of factors. This factorization of the joint probability is represented as the factor graph depicted in Fig. 3. The factor graph representation is central in the formulation of the receiver algorithm. In the following we introduce the variables and factors in the factor graph, moving from right to left.[9]

The interleaving, coding and modulation of the data bits are described in Sec. II-A. The subgraph characterizing the system implementing these tasks involves the factors

$$\begin{aligned} f_{u_k}(u_k) &\triangleq p(u_k) = 0.5\,\mathbb{1}_{[u_k\in\{0,1\}]}, && k \in \mathcal{K}, \\ f_C(\boldsymbol{c},\boldsymbol{u}) &\triangleq p(\boldsymbol{c}|\boldsymbol{u}) = \mathbb{1}_{[\boldsymbol{c}=\mathcal{C}(\boldsymbol{u})]}, \\ f_{M_i}(x_i,\boldsymbol{c}^{(i)}) &\triangleq p(x_i|\boldsymbol{c}^{(i)}) = \mathbb{1}_{[x_i=\mathcal{M}(\boldsymbol{c}^{(i)})]}, && i \in \mathcal{D}, \end{aligned}$$

where $\mathcal{K} = \{1,\dots,K\}$ is the index set of the information bits. The factor $f_C(\boldsymbol{c},\boldsymbol{u})$ describes the interleaving and channel coding processes. By "zooming in" this factor can be expanded to a subgraph involving auxiliary variables and factors that describe the structure of the channel code and interleaver.

The subgraph characterizing the observation process described by (6) and (11) involves the following factors for pilot- and data subcarriers, respectively:

$$\begin{aligned} f_{P_j}(\boldsymbol{\alpha},\boldsymbol{\tau},\beta) &\triangleq p(y_j|\boldsymbol{\alpha},\boldsymbol{\tau};\beta) \\ &= \mathrm{CN}(y_j; x_j[\boldsymbol{\Psi}(\boldsymbol{\tau})\boldsymbol{\alpha}]_j, \beta), && j \in \mathcal{P}, \\ f_{D_i}(x_i,\boldsymbol{\alpha},\boldsymbol{\tau},\beta) &\triangleq p(y_i|x_i,\boldsymbol{\alpha},\boldsymbol{\tau};\beta) \\ &= \mathrm{CN}(y_i; x_i[\boldsymbol{\Psi}(\boldsymbol{\tau})\boldsymbol{\alpha}]_i, \beta), && i \in \mathcal{D}. \end{aligned}$$

The lth virtual multipath components is modelled through the variables α_l, τ_l and z_l. To ease the terminology we drop the attribute "virtual" in the following. To model the fact that there are only a multipath components, a Bernoulli-Gaussian prior is used. This prior assigns large probability to the event $\alpha_l = 0$. The model contains $L_{\max}$ multipath components of which only a subset is activated, i.e. has $\alpha_l \neq 0$. The number $L_{\max}$ is an upper bound on the number of multipath components that can be estimated.[10] This allows us to derive an algorithm that inherently estimates the number of multipath components. Each component is assigned an activation variable $z_l \in \{0,1\}$, which is 1 when said multipath component is active and 0 otherwise. The sequence $\{z_1,\dots,z_{L_{\max}}\}$ is modelled i.i.d. where each z_l is assigned a Bernoulli prior with activation probability ρ:

$$f_{z_l}(z_l,\rho) \triangleq p(z_l;\rho) = \mathrm{Bern}(z_l;\rho), \qquad l \in \mathcal{L}, \tag{12}$$

where we have defined the set of multipath component indices $\mathcal{L} = \{1,\dots,L_{\max}\}$. The prior density of the multipath coefficient α_l is conditioned on z_l, such that $z_l = 0$ implies $\alpha_l = 0$ and $z_l = 1$ gives a Gaussian density with variance η:

$$\begin{aligned} f_{\alpha_l}(\alpha_l,z_l,\eta) &\triangleq p(\alpha_l|z_l;\eta) \\ &= (1-z_l)\delta(\alpha_l) + z_l\,\mathrm{CN}(\alpha_l;0,\eta), \quad l \in \mathcal{L}. \end{aligned}$$

When performing inference in this model, the estimated number of active multipath components is $\hat{L} \triangleq \|\hat{\boldsymbol{\alpha}}\|_0$, where $\hat{\boldsymbol{\alpha}}$ is a vector containing the estimates of α_l for all $l \in \mathcal{L}$.

We finally need to impose a prior model on the multipath delays τ_l, $l \in \mathcal{L}$. The only prior information available is through the assumption (3) that implies that for all $l \in \mathcal{L}$ we have $0 \leq \tau_l \leq T_{\mathrm{CP}}$. To express this an i.i.d. uniform prior is used:

$$f_{\tau_l}(\tau_l) \triangleq p(\tau_l) = \mathrm{unif}(\tau_l;0,T_{\mathrm{CP}}), \qquad l \in \mathcal{L}.$$

III. Inference Method

The BER optimal receiver (assuming ρ, η and β known) computes the MAP estimate

$$\hat{u}_k = \underset{u_k\in\{0,1\}}{\arg\max}\; p(u_k|\boldsymbol{y};\rho,\eta,\beta), \quad k \in \mathcal{K}. \tag{13}$$

The pdf $p(u_k|\boldsymbol{y};\rho,\eta,\beta) \propto p(u_k,\boldsymbol{y};\rho,\eta,\beta)$ can ideally be found by marginalizing all variables but u_k in the joint pdf

$$p(\boldsymbol{y},\boldsymbol{z},\boldsymbol{\alpha},\boldsymbol{\tau},\boldsymbol{x}_{\mathcal{D}},\boldsymbol{c},\boldsymbol{u};\rho,\eta,\beta) = p(\boldsymbol{y}|\boldsymbol{x}_{\mathcal{D}},\boldsymbol{\alpha},\boldsymbol{\tau};\beta)$$

[9]We abuse terminology and associate variables and factors with their respective nodes in the factor graph.

[10]In our implementation we select $L_{\max} = \lceil T_{\mathrm{CP}} N \Delta_f \rceil + 1$, which is the maximum number of degrees of freedom under the assumption (3) [18]. It roughly corresponds to the number of baud-spaced (spacing $1/(N\Delta_f)$) components on the interval $[0, T_{\mathrm{CP}})$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2868314, IEEE Transactions on Signal Processing

$$\prod_{l\in\mathcal{L}} p(\alpha_l|z_l;\eta)p(z_l;\rho)p(\tau_l)\prod_{i\in\mathcal{D}} p(x_i|\boldsymbol{c}^{(i)})p(\boldsymbol{c}|\boldsymbol{u})\prod_{k\in\mathcal{K}} p(u_k).$$

Calculating the marginals of u_k, $k\in\mathcal{K}$, is intractable and we resort to approximate Bayesian inference.

A. Combined Belief Propagation and Mean-Field

Our inference method is based on the merged belief propagation and mean-field (BP-MF) framework of [8]. In this framework a so-called belief function is found for each variable in the factor graph. The belief function is an approximation of the marginal posterior pdf or pmf of that variable. We abuse notation and let $q(a)$ denote the belief of variable a. When the set of belief functions has been calculated, the MAP estimate of the kth data bit is found as the mode of $q(u_k)$.

For tractability we obtain a point estimate of the variables $\boldsymbol{z}$ and $\boldsymbol{\tau}$. This is achieved as proposed in [8] by restricting their beliefs to be Kronecker and Dirac delta functions, i.e., $q(z_l)=\delta[z_l-\hat{z}_l]$ and $q(\tau_l)=\delta(\tau_l-\hat{\tau}_l)$ for all $l\in\mathcal{L}$.

At the heart of BP-MF lies the so-called region-based free energy approximation (RBFE) [51]. The RBFE is obtained by splitting the factor graph into a MF and a BP subgraph, as indicated in Fig. 3. The RBFE is a function of[11] the point estimates $\hat{\boldsymbol{z}},\hat{\boldsymbol{\tau}}$ and the belief functions $q(\alpha_l),q(x_i),q([\boldsymbol{c}^{(i)}]_m)$ and $q(u_k)$ for indices $l\in\mathcal{L},i\in\mathcal{D},k\in\mathcal{K}$ and $m=1,\dots,Q$. It is also a function of the model parameter estimates $(\hat{\rho},\hat{\eta},\hat{\beta})$, as justified below. The expression of the RBFE is given in Appendix A. BP-MF seeks to minimize the RBFE under a number of normalization and consistency constraints. The messages of BP-MF are derived such that at convergence they satisfy the Karush-Kuhn-Tucker conditions of the constrained RBFE minimization, i.e., a (possibly local) minimum of the constrained problem is found. See [8] for a more detailed discussion of BP-MF.

The understanding of BP-MF as RBFE minimization allows us to make a number of adaptations to the message-passing scheme to improve convergence speed. Further, we will see that this understanding is useful when analyzing convergence of the algorithm.

B. Model Parameter Estimation with BP-MF

The BP-MF framework [8] does not directly provide a method to estimate the unknown model parameters (ρ,η,β). We propose to do so by letting the RBFE be a function of these model parameters. The model parameter estimates $(\hat{\rho},\hat{\eta},\hat{\beta})$ are then obtained as the minimizers of the RBFE.

To justify this method we first note that the model parameters are located in the MF subgraph. Then we follow an approach similar to [52] to obtain a lower bound on the marginal log-likelihood function:

$$\ln p(\boldsymbol{y};\hat{\rho},\hat{\eta},\hat{\beta})\geq -F_{\text{BP-MF}}+\text{const.}, \tag{14}$$

where $F_{\text{BP-MF}}$ is the RBFE (39) and the constant only depends on beliefs of variables in the BP subgraph (including $q(x_i)$, for $i\in\mathcal{D}$), i.e., it does not depend on $(\hat{\rho},\hat{\eta},\hat{\beta})$. It can then be seen that the values of $(\hat{\rho},\hat{\eta},\hat{\beta})$ minimizing $F_{\text{BP-MF}}$ maximize the lower bound on the likelihood function in (14). These minimizers are thus approximate ML estimates. We note that if the above approach is applied in a pure MF context it simplifies to variational EM estimation with all other variables treated as latent variables [8], [53].

[11]The RBFE is also a functional of the beliefs corresponding to the factors in the BP subgraph. BP-MF enforces consistency between the variable beliefs and these factor beliefs. Since the latter are not relevant to the derivation of the receiver, we omit them.

C. Relation to Prior Art

To relate our receiver algorithm to current methods we note that the decoding of many popular channel codes can be described as an instance of BP [54] in a factor graph [55]–[57]. For example, BP decoding of a convolutional code leads to the BCJR algorithm [58]. We see in Fig. 3 that the merged BP-MF algorithm employs BP in the subgraph that represents the channel code, i.e., standard techniques are used for decoding.

Similarly, there are examples in the literature of MF inference where the underlying factor graph resembles the MF subgraph of our receiver. The work [31] uses a Bernoulli-Gaussian prior model similar to that in our work, while [30], [32] use a gamma-Gaussian prior typical of sparse Bayesian learning.

The strength of the BP-MF framework is now clear: It allows us to merge existing methods for channel decoding and sparsity-based estimation using a unifying design method (namely that of RBFE minimization).

IV. Parametric BP-MF Receiver

To minimize the RBFE, we apply the BP-MF algorithm given by Eq. (21)–(22) in [8] on the factor graph of Fig. 3. In the following we use the notation $\langle\cdot\rangle_a$ to denote expectation with respect to the belief density $q(a)$. We follow the convention of [8] in naming the messages. In [9] a similar BP-MF receiver is derived, which does not exploit channel sparsity.

A. Message Passing for Channel Estimation

1) Update of Coefficient Belief: We start by finding belief updates in the MF subgraph. To find the update of $q(\alpha_l)$, $l\in\mathcal{L}$, we calculate the messages passed to the node α_l:

$$m^{\text{MF}}_{f_{\alpha_l}\to\alpha_l}(\alpha_l)\propto\begin{cases}\exp\left(-\hat{\eta}^{-1}|\alpha_l|^2\right) & \text{if } \hat{z}_l=1,\\ \delta(\alpha_l) & \text{if } \hat{z}_l=0\end{cases}$$

$$m^{\text{MF}}_{f_{D_i}\to\alpha_l}(\alpha_l)\propto\exp\Big(-\hat{\beta}^{-1}\left\langle|y_i-x_i[\boldsymbol{\Psi}(\hat{\boldsymbol{\tau}})\boldsymbol{\alpha}]_i|^2\right\rangle_{x_i,\boldsymbol{\alpha}_{\setminus l}}\Big)$$

$$m^{\text{MF}}_{f_{P_j}\to\alpha_l}(\alpha_l)\propto\exp\Big(-\hat{\beta}^{-1}\left\langle|y_j-x_j[\boldsymbol{\Psi}(\hat{\boldsymbol{\tau}})\boldsymbol{\alpha}]_j|^2\right\rangle_{\boldsymbol{\alpha}_{\setminus l}}\Big),$$

which holds for all $l\in\mathcal{L}$, $i\in\mathcal{D}$ and $j\in\mathcal{P}$. Taking the product of all messages going into the node α_l gives its belief

$$q(\alpha_l)=\begin{cases}\text{CN}(\alpha_l;\hat{\mu}_l,\hat{\sigma}_l^2) & \text{if } \hat{z}_l=1\\ \delta(\alpha_l) & \text{if } \hat{z}_l=0\end{cases} \tag{15}$$

with the active component mean and variance

$$\hat{\mu}_l=\hat{\sigma}_l^2 q_l \tag{16}$$

$$\hat{\sigma}_l^2=\left(s_l+\hat{\eta}^{-1}\right)^{-1}. \tag{17}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2868314, IEEE Transactions on Signal Processing

Here we have introduced

$$s_l = \hat{\beta}^{-1}\boldsymbol{\psi}^{\mathrm{H}}(\hat{\tau}_l)\left\langle \mathbf{X}^{\mathrm{H}}\mathbf{X}\right\rangle_{\boldsymbol{x}_{\mathcal{D}}}\boldsymbol{\psi}(\hat{\tau}_l) \tag{18}$$

$$q_l = \hat{\beta}^{-1}\boldsymbol{\psi}^{\mathrm{H}}(\hat{\tau}_l)\boldsymbol{r} \tag{19}$$

$$\boldsymbol{r} = \langle \mathbf{X}\rangle^{\mathrm{H}}_{\boldsymbol{x}_{\mathcal{D}}}\boldsymbol{y} - \left\langle \mathbf{X}^{\mathrm{H}}\mathbf{X}\right\rangle_{\boldsymbol{x}_{\mathcal{D}}}\boldsymbol{\Psi}(\hat{\boldsymbol{\tau}}_{\hat{\mathcal{A}}\setminus l})\hat{\boldsymbol{\mu}}_{\hat{\mathcal{A}}\setminus l}, \tag{20}$$

where $\boldsymbol{\psi}(\tau_l)$ is defined as the lth column of $\boldsymbol{\Psi}(\boldsymbol{\tau})$. Note that the belief of inactive components ($\hat{z}_l = 0$) becomes a point mass at $\alpha_l = 0$, thus eliminating the influence of that component in the product $\mathbf{X}\boldsymbol{\Psi}(\hat{\boldsymbol{\tau}})\boldsymbol{\alpha}$. We have defined the set of currently active components as $\hat{\mathcal{A}} \triangleq \{l : \hat{z}_l = 1\}$ and the vectors $\hat{\boldsymbol{\mu}} = [\hat{\mu}_1, \ldots, \hat{\mu}_{L_{\max}}]^{\mathrm{T}}$, $\hat{\boldsymbol{\sigma}}^2 = [\hat{\sigma}_1^2, \ldots, \hat{\sigma}_{L_{\max}}^2]^{\mathrm{T}}$.

2) Joint Update of Delay and Coefficient Belief: We now turn our attention to the estimation of the multipath delays τ_l, $l \in \mathcal{L}$. To improve the convergence speed of the algorithm, we find the update of $\hat{\tau}_l$ by minimizing the RBFE *jointly* with respect to the beliefs $q(\alpha_l)$ and $\hat{\tau}_l$. Due to the selected prior $p(\tau_l)$, the following expressions are valid for $\hat{\tau}_l \in [0, T_{\mathrm{CP}}]$. We are only concerned with active components, i.e., $l \in \hat{\mathcal{A}}$ and thus $\hat{z}_l = 1$. Writing only the terms of the RBFE (39) that depend on $q(\alpha_l)$ and $\hat{\tau}_l$, we get

$$F_{\text{BP-MF}}(q(\alpha_l), \hat{\tau}_l) \propto^{\mathrm{e}} \int q(\alpha_l)\ln\frac{q(\alpha_l)}{Q(\alpha_l, \hat{\tau}_l)}\,\mathrm{d}\alpha_l \tag{21}$$

with

$$\begin{aligned} Q(\alpha_l, \hat{\tau}_l) &= p(\alpha_l|\hat{z}_l;\hat{\eta})p(\hat{\tau}_l)\exp\left(\left\langle \ln p(\boldsymbol{y}|\boldsymbol{x}_{\mathcal{D}}, \boldsymbol{\alpha}, \hat{\boldsymbol{\tau}}; \hat{\beta})\right\rangle_{\boldsymbol{x}_{\mathcal{D}}, \boldsymbol{\alpha}_{\setminus l}}\right) \\ &\propto \mathrm{CN}(\alpha_l; \hat{\mu}_l, \hat{\sigma}_l^2)\exp\left(\frac{|q_l|^2}{s_l + \hat{\eta}^{-1}}\right), \end{aligned} \tag{22}$$

where $\hat{\sigma}_l^2$, $\hat{\mu}_l$, s_l and q_l are given by (16) - (19) and thus implicitly are functions of $\hat{\tau}_l$. We need to minimize (21) under the normalization constraint $\int q(\alpha_l)\,\mathrm{d}\alpha_l = 1$. To do so, define

$$g_{\tau_l}(\hat{\tau}_l) \triangleq \max_{\tilde{q}(\alpha_l):\int \tilde{q}(\alpha_l)\,\mathrm{d}\alpha_l = 1} -F_{\text{BP-MF}}\left(\tilde{q}(\alpha_l), \hat{\tau}_l\right) \tag{23}$$

$$\propto^{\mathrm{e}} \ln\int Q(\alpha_l, \hat{\tau}_l)\,\mathrm{d}\alpha_l \tag{24}$$

$$\propto^{\mathrm{e}} \frac{\hat{\beta}^{-2}}{s_l + \hat{\eta}^{-1}}|\boldsymbol{\psi}^{\mathrm{H}}(\hat{\tau}_l)\boldsymbol{r}|^2. \tag{25}$$

The result in (24) is easily obtained by noting that (21) can be rewritten as

$$F_{\text{BP-MF}} \propto^{\mathrm{e}} \mathrm{KL}\left[q(\alpha_l)\middle\|\frac{Q(\alpha_l, \hat{\tau}_l)}{\int Q(\tilde{\alpha}_l, \hat{\tau}_l)\,\mathrm{d}\tilde{\alpha}_l}\right] - \ln\int Q(\tilde{\alpha}_l, \hat{\tau}_l)\,\mathrm{d}\tilde{\alpha}_l,$$

where $\mathrm{KL}[\cdot||\cdot]$ is the Kullback-Leibler divergence. The coefficient belief is selected as the maximizer of (23), i.e., $q(\alpha_l) = Q(\alpha_l, \hat{\tau}_l)/\int Q(\tilde{\alpha}_l, \hat{\tau}_l)\,\mathrm{d}\tilde{\alpha}_l$, which is easily shown to coincide with the result in (15).

Since s_l is constant with respect to $\hat{\tau}_l$, we find the delay update as

$$\hat{\tau}_l = \underset{\tilde{\tau}_l \in [0, T_{\mathrm{CP}}]}{\arg\max}\ g_{\tau_l}(\tilde{\tau}_l) = \underset{\tilde{\tau}_l \in [0, T_{\mathrm{CP}}]}{\arg\max}\ |\boldsymbol{\psi}^{\mathrm{H}}(\tilde{\tau}_l)\boldsymbol{r}|^2. \tag{26}$$

We recognize the objective function in (26) as the continuous periodogram of the residual vector $\boldsymbol{r}$. While it is possible to find the maximizer of the periodogram, doing so has high computational cost. In our iterative algorithm, we instead find an update of $\hat{\tau}_l$ that cannot increase the objective in (26). Denote the updated delay estimate as $\hat{\tau}_l^{[t]}$ and the previous delay estimate as $\hat{\tau}_l^{[t-1]}$. Our scheme now reads:

1) Find initial step $\Delta = \frac{g'_\tau(\hat{\tau}_l^{[t-1]})}{|g''_\tau(\hat{\tau}_l^{[t-1]})|}$.
2) If $g_\tau(\hat{\tau}_l^{[t-1]} + \Delta) \geq g_\tau(\hat{\tau}_l^{[t-1]})$, set $\hat{\tau}_l^{[t]} = \hat{\tau}_l^{[t-1]} + \Delta$ and terminate. Otherwise set $\Delta = \frac{\Delta}{2}$ and repeat step 2.

Functions $g'_\tau(\tau_l)$ and $g''_\tau(\tau_l)$ are the first and second derivatives of (25). The scheme gives the Newton update of $\hat{\tau}_l$ if this value increases the objective function and otherwise resorts to a gradient ascent with a backtracking line search. We have the following lemma, that we will use in the convergence analysis:

Lemma 1: The procedure listed in Steps 1-2 above followed by an update of $q(\alpha_l)$ does not increasing the RBFE.

Proof: First, note that the updated $\hat{\tau}_l$, does not decrease $g_{\tau_l}(\hat{\tau}_l)$. It then follows that by selecting the maximizer of (23), the RBFE is non-increasing. ∎

3) Joint Update of Activation Variable and Coefficient Belief: We now turn our focus on the update of the activation variable $\hat{z}_l$. It is again desirable to perform a joint update of $\hat{z}_l$ and $q(\alpha_l)$. We proceed in a similar way as we did to compute the updates of the multipath delays. The terms in the RBFE (39), which depend on $q(\alpha_l)$ and $\hat{z}_l$, are denoted as $F_{\text{BP-MF}}(q(\alpha_l), \hat{z}_l)$. We then define

$$g_{z_l}(\hat{z}_l) \triangleq \max_{\tilde{q}(\alpha_l):\int \tilde{q}(\alpha_l)\,\mathrm{d}\alpha_l = 1} -F_{\text{BP-MF}}(\tilde{q}(\alpha_l), \hat{z}_l) \tag{27}$$

$$\propto^{\mathrm{e}} \begin{cases} \frac{|\hat{\mu}_l|^2}{\hat{\sigma}_l^2} + \ln\frac{\hat{\sigma}_l^2}{\hat{\eta}} + \ln\hat{\rho} & \text{if } \hat{z}_l = 1, \\ \ln(1-\hat{\rho}) & \text{if } \hat{z}_l = 0. \end{cases} \tag{28}$$

This result is easily obtained by following steps analogous to (21) – (25). The activation variable solves the decision problem $\hat{z}_l = \max_{\tilde{z}_l \in \{0,1\}} g_{z_l}(\tilde{z}_l)$. Writing the "activation criterion" $g_{z_l}(1) > g_{z_l}(0)$ we get

$$\frac{|\hat{\mu}_l|^2}{\hat{\sigma}_l^2} > \ln\frac{\hat{\eta}}{\hat{\sigma}_l^2} + \ln\frac{1-\hat{\rho}}{\hat{\rho}}. \tag{29}$$

If the above criterion is true we set $\hat{z}_l = 1$; otherwise we set $\hat{z}_l = 0$. The corresponding update of $q(\alpha_l)$ is the maximizer of (27), which remains as in (15). The criterion in (29) is the same as that obtained in [31].

4) Update of Channel Parameter Estimates: The channel parameters (ρ, η, β) are estimated as the values that minimize the RBFE. Writing only the terms of the RBFE (39) that depend on the channel parameters we have

$$\begin{aligned} &F_{\text{BP-MF}}(\hat{\rho}, \hat{\eta}, \hat{\beta}) \\ &\quad\propto^{\mathrm{e}} -\left\langle \ln\prod_{l\in\mathcal{L}} p(\hat{z}_l;\hat{\rho})p(\alpha_l|\hat{z}_l;\hat{\eta})p(\boldsymbol{y}|\boldsymbol{\alpha}, \hat{\boldsymbol{\tau}}, \boldsymbol{x}_{\mathcal{D}}; \hat{\beta})\right\rangle_{\boldsymbol{x}_{\mathcal{D}}, \boldsymbol{\alpha}} \\ &\quad\propto^{\mathrm{e}} \|\hat{\boldsymbol{z}}\|_0 \ln\hat{\rho} + (L_{\max} - \|\hat{\boldsymbol{z}}\|_0)\ln(1-\hat{\rho}) - N\ln\hat{\beta} - \hat{\beta}^{-1}u \\ &\qquad - \|\hat{\boldsymbol{z}}\|_0 \ln\hat{\eta} - \hat{\eta}^{-1}\sum_{\{l:\hat{z}_l=1\}}(|\hat{\mu}_l|^2 + \hat{\sigma}_l^2), \end{aligned} \tag{30}$$

where

$$\begin{aligned} u &\triangleq \left\langle \|\boldsymbol{y} - \mathbf{X}\boldsymbol{\Psi}(\hat{\boldsymbol{\tau}})\boldsymbol{\alpha}\|^2\right\rangle_{\boldsymbol{x}_{\mathcal{D}}, \boldsymbol{\alpha}} \\ &= \|\boldsymbol{y}\|^2 + \hat{\boldsymbol{\mu}}_{\hat{\mathcal{A}}}^{\mathrm{H}}\boldsymbol{\Psi}^{\mathrm{H}}(\hat{\boldsymbol{\tau}}_{\hat{\mathcal{A}}})\left\langle \mathbf{X}^{\mathrm{H}}\mathbf{X}\right\rangle_{\boldsymbol{x}_{\mathcal{D}}}\boldsymbol{\Psi}(\hat{\boldsymbol{\tau}}_{\hat{\mathcal{A}}})\hat{\boldsymbol{\mu}}_{\hat{\mathcal{A}}} \end{aligned}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2868314, IEEE Transactions on Signal Processing

$$+\sum_{l\in\hat{\mathcal{A}}}\hat{\sigma}_l^2\boldsymbol{\psi}^{\mathrm{H}}(\hat{\tau}_l)\left\langle\mathbf{X}^{\mathrm{H}}\mathbf{X}\right\rangle_{\boldsymbol{x}_{\mathcal{D}}}\boldsymbol{\psi}(\hat{\tau}_l) -2\,\mathrm{Re}\left\{\boldsymbol{y}^{\mathrm{H}}\left\langle\mathbf{X}\right\rangle_{\boldsymbol{x}_{\mathcal{D}}}\boldsymbol{\Psi}(\hat{\boldsymbol{\tau}}_{\hat{\mathcal{A}}})\hat{\boldsymbol{\mu}}_{\hat{\mathcal{A}}}\right\}. \quad (31)$$

It is readily seen that $F_{\text{BP-MF}}(\hat{\rho},\hat{\eta},\hat{\beta})$ can be minimized independently with respect to each of the parameters. By taking derivatives and equating to zero we find the global minima (the second derivatives are all positive):

$$\hat{\rho}=\frac{\|\hat{\boldsymbol{z}}\|_0}{L_{\max}} \quad (32)$$

$$\hat{\eta}=\frac{\sum_{\{l:\hat{z}_l=1\}}(|\hat{\mu}_l|^2+\hat{\sigma}_l^2)}{\|\hat{\boldsymbol{z}}\|_0} \quad (33)$$

$$\hat{\beta}=\frac{u}{N}. \quad (34)$$

5) Iterating all Coefficient Beliefs Ad-Infinitum: In [32] it is demonstrated that iterating the updates of some variables *ad-infinitum* is a powerful technique for increasing the convergence speed of MF algorithms. We apply that idea to the beliefs of the multipath coefficients.

Since $q(\alpha_l)=\delta(\alpha_l)$ for all $l\in\mathcal{L}\backslash\hat{\mathcal{A}}$, the following discussion is only concerned with the beliefs of active components, i.e. for $l\in\hat{\mathcal{A}}$. First note that the variance (17) of an active multipath coefficient $\hat{\sigma}_l^2$ does not depend on the beliefs of the remaining coefficients $q(\alpha_k)$, $k\neq l$. The mean (16) of the lth coefficient, on the other hand, depends on the remaining mean values as

$$\hat{\mu}_l=\underbrace{\hat{\sigma}_l^2}_{[\mathbf{Q}]_{l,l}^{-1}}\left(\underbrace{\hat{\beta}^{-1}\boldsymbol{\psi}^{\mathrm{H}}(\hat{\tau}_l)\langle\mathbf{X}\rangle_{\boldsymbol{x}_{\mathcal{D}}}^{\mathrm{H}}\boldsymbol{y}}_{p_l} -\sum_{k\in\hat{\mathcal{A}}\backslash l}\underbrace{\hat{\beta}^{-1}\boldsymbol{\psi}^{\mathrm{H}}(\hat{\tau}_l)\langle\mathbf{X}^{\mathrm{H}}\mathbf{X}\rangle_{\boldsymbol{x}_{\mathcal{D}}}\boldsymbol{\psi}(\hat{\tau}_k)}_{[\mathbf{Q}]_{l,k}}\hat{\mu}_k\right)$$

for all $l\in\hat{\mathcal{A}}$. The matrix $\mathbf{Q}$ is of size $|\hat{\mathcal{A}}|\times|\hat{\mathcal{A}}|$ and we have abused notation in using l,k as indices into this matrix, because $1\leq l,k\leq L_{\max}$, even though $|\hat{\mathcal{A}}|\leq L_{\max}$. The above equation is recognized as the Gauss-Seidel [59] iteration for solving the system of linear equations

$$\mathbf{Q}\hat{\boldsymbol{\mu}}_{\hat{\mathcal{A}}}=\boldsymbol{p} \quad (35)$$

with

$$\boldsymbol{p}=\hat{\beta}^{-1}\boldsymbol{\Psi}^{\mathrm{H}}(\hat{\boldsymbol{\tau}}_{\hat{\mathcal{A}}})\left\langle\mathbf{X}\right\rangle_{\boldsymbol{x}_{\mathcal{D}}}^{\mathrm{H}}\boldsymbol{y}$$
$$\mathbf{Q}=\hat{\beta}^{-1}\boldsymbol{\Psi}^{\mathrm{H}}(\hat{\boldsymbol{\tau}}_{\hat{\mathcal{A}}})\left\langle\mathbf{X}^{\mathrm{H}}\mathbf{X}\right\rangle_{\boldsymbol{x}_{\mathcal{D}}}\boldsymbol{\Psi}(\hat{\boldsymbol{\tau}}_{\hat{\mathcal{A}}})+\hat{\eta}^{-1}\mathbf{I}.$$

It follows that the updates of $\hat{\mu}_l$, for all $l\in\hat{\mathcal{A}}$, converge to the solution $\hat{\boldsymbol{\mu}}_{\hat{\mathcal{A}}}$ found by solving (35).

We note that in the hypothetical special case where the beliefs of $\mathbf{X}$ are point estimates (or equivalently known) $\boldsymbol{y}=\mathbf{X}\boldsymbol{\Psi}(\hat{\boldsymbol{\tau}}_{\hat{\mathcal{A}}})\boldsymbol{\alpha}_{\hat{\mathcal{A}}}+\boldsymbol{w}$ is a linear observation model with Gaussian noise. In this case, the estimator $\hat{\boldsymbol{\mu}}_{\hat{\mathcal{A}}}=\mathbf{Q}^{-1}\boldsymbol{p}$ reduces to the LMMSE estimator of $\boldsymbol{\alpha}_{\hat{\mathcal{A}}}$ in the linear observation model under the Bayesian model dictated by the current beliefs of the remaining variables. The estimator $\hat{\boldsymbol{\mu}}_{\hat{\mathcal{A}}}=\mathbf{Q}^{-1}\boldsymbol{p}$ is, however, not the LMMSE estimator of $\boldsymbol{\alpha}_{\hat{\mathcal{A}}}$ when the uncertainty of the estimate of $\mathbf{X}$ is considered.

B. Message-Passing for Decoding

In the previous subsections we derived the *belief functions* $q(\cdot)$ of the variables whose factor neighbours are in the MF subgraph only. To perform inference in the BP subgraph, i.e., detection, demapping, decoding and deinterleaving, we need to calculate the *messages* that are passed along its edges.

We begin with the messages $n_{x_i\to f_{M_i}}(x_i)$, $i\in\mathcal{D}$, which constitute the interface from the continuous-valued channel estimator to the discrete-valued decoder. They are given as

$$n_{x_i\to f_{M_i}}(x_i)=m^{\text{MF}}_{f_{D_i}\to x_i}(x_i) \propto\mathrm{CN}\left(x_i;\frac{y_i\left\langle g_i\right\rangle^*_{\boldsymbol{\alpha},\boldsymbol{\tau}}}{\left\langle|g_i|^2\right\rangle_{\boldsymbol{\alpha},\boldsymbol{\tau}}},\frac{\hat{\beta}}{\left\langle|g_i|^2\right\rangle_{\boldsymbol{\alpha},\boldsymbol{\tau}}}\right), \quad (36)$$

where $g_i\triangleq[\boldsymbol{\Psi}(\boldsymbol{\tau})\boldsymbol{\alpha}]_i$ is the CFR sampled at subcarrier i. Its mean and second moment are

$$\left\langle g_i\right\rangle_{\boldsymbol{\alpha},\boldsymbol{\tau}}=[\boldsymbol{\Psi}(\hat{\boldsymbol{\tau}})\hat{\boldsymbol{\mu}}]_i$$
$$\left\langle|g_i|^2\right\rangle_{\boldsymbol{\alpha},\boldsymbol{\tau}}=\left[\boldsymbol{\Psi}(\hat{\boldsymbol{\tau}})(\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^{\mathrm{H}}+\mathrm{diag}(\hat{\boldsymbol{\sigma}}^2))\boldsymbol{\Psi}^{\mathrm{H}}(\hat{\boldsymbol{\tau}})\right]_{i,i}.$$

Note that even though the above expression has the form of a Gaussian, the messages are probability mass functions obtained by evaluating the above Gaussian at the points of the symbol alphabet $\mathbb{A}_{\mathrm{D}}$ followed by appropriate normalization.

The mean in (36) can be interpreted as the output of an LMMSE equalizer. Consider the observation model $y_i=g_ix_i+w_i$ where $p(w_i)=\mathrm{CN}(w_i;0,\hat{\beta})$ and $g_i=[\boldsymbol{\Psi}(\hat{\boldsymbol{\tau}})\boldsymbol{\alpha}]_i$. Let $q(\alpha_l)$ be the density of α_l and impose a prior $p(x_i)=\mathrm{CN}(x_i;0,\sigma_{x_i}^2)$ on x_i. The LMMSE estimator of x_i is now

$$\hat{x}_i^{\text{LMMSE}}=\frac{y_i\left\langle g_i\right\rangle^*_{\boldsymbol{\alpha},\boldsymbol{\tau}}}{\left\langle|g_i|^2\right\rangle_{\boldsymbol{\alpha},\boldsymbol{\tau}}+\hat{\beta}\sigma_{x_i}^{-2}}.$$

By letting $\sigma_{x_i}^2\to\infty$ to express that we have no prior information on x_i, we recover the mean in (36). Note that a similar analogy does not exist for the variance in (36).

All remaining messages passed in the BP subgraph are functions of discrete variables (i.e., coded or information bits). These messages are calculated with the sum-product algorithm, see e.g. [55], [56]. Due to space constraints, we do not give the details here.

When BP messages have been passed in the BP subgraph, the beliefs of the data symbols x_i, $i\in\mathcal{D}$, are calculated from

$$q(x_i)\propto m^{\text{MF}}_{f_{D_i}\to x_i}(x_i)m^{\text{BP}}_{f_{M_i}\to x_i}(x_i). \quad (37)$$

Since $q(x_i)$ is a probability mass function, we can use straightforward evaluation of finite sums to obtain $\left\langle\mathbf{X}\right\rangle_{\boldsymbol{x}_{\mathcal{D}}}$ and $\left\langle\mathbf{X}^{\mathrm{H}}\mathbf{X}\right\rangle_{\boldsymbol{x}_{\mathcal{D}}}$, which are used in the belief updates in the MF subgraph.

C. An Incremental Algorithm

Algorithm 1 combines the derived belief update expressions into an iterative receiver with sparsity-based parametric channel estimation. The algorithm is split into two parts: channel estimation (lines 5 - 30) and decoding (line 32). The outer

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2868314, IEEE Transactions on Signal Processing

Algorithm 1: Parametric BP-MF receiver.

Input: Observations $\boldsymbol{y}$, pilot indices $\mathcal{P}$ and pilot symbols $\boldsymbol{x}_{\mathcal{P}}$.
Output: Belief functions of data bits $\{q(u_k)\}_{k\in\mathcal{K}}$.
Notes: Define the set of components as $\mathcal{L} = \{1, \ldots, L_{\max}\}$ and the set of active components as $\hat{\mathcal{A}} \triangleq \{l \in \mathcal{L} : \hat{z}_l = 1\}$.

1 $\tilde{\boldsymbol{\tau}} \leftarrow$ Vector with values from equispaced grid on $[-\frac{1/2}{N\Delta_f}, T_{\text{CP}}]$.
2 Initialize channel parameter estimates $(\hat{\rho}, \hat{\eta}, \hat{\beta})$.
3 $\hat{\boldsymbol{z}}, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}^2 \leftarrow$ Zero vectors of length N.
4 **while** *Outer stopping criterion not met* **do**
5 **while** *Inner stopping criterion not met* **do**
6 $\hat{\boldsymbol{\mu}}_{\hat{\mathcal{A}}}, \hat{\boldsymbol{\sigma}}^2_{\hat{\mathcal{A}}} \leftarrow$ Updates from (35) and (17).
7 *Activate an inactive component:*
8 **if** *the inactive set* $\mathcal{L}\backslash\hat{\mathcal{A}}$ *is non-empty* **then**
9 $l \leftarrow$ Any index from the inactive set $\mathcal{L}\backslash\hat{\mathcal{A}}$.
10 $\hat{z}_l \leftarrow 1$.
11 $\hat{\tau}_l \leftarrow$ Value from (26) calculated on the grid $\tilde{\boldsymbol{\tau}}$.
12 $\hat{\boldsymbol{\mu}}_{\hat{\mathcal{A}}}, \hat{\boldsymbol{\sigma}}^2_{\hat{\mathcal{A}}} \leftarrow$ Updates from (35) and (17).
13 $\hat{\tau}_l \leftarrow$ Update via the scheme in Sec. IV-A2.
14 $\hat{\mu}_l, \hat{\sigma}^2_l \leftarrow$ Updates from (16) and (17).
15 **if** *activation criterion* (29) *is false* **then**
16 $\hat{z}_l \leftarrow 0$.
17 Reset $\hat{\boldsymbol{\mu}}_{\hat{\mathcal{A}}}$ to the value calculated in line 6.
18 **end**
19 **end**
20 *Update all components currently included in model:*
21 **for** $l \in \hat{\mathcal{A}}$ **do**
22 $\hat{\tau}_l \leftarrow$ Update via the scheme in Sec. IV-A2.
23 $\hat{\mu}_l, \hat{\sigma}^2_l \leftarrow$ Updates from (16) and (17).
24 **if** *activation criterion* (29) *is false* **then**
25 $\hat{z}_l \leftarrow 0$.
26 **end**
27 **end**
28 $\hat{\boldsymbol{\mu}}_{\hat{\mathcal{A}}}, \hat{\boldsymbol{\sigma}}^2_{\hat{\mathcal{A}}} \leftarrow$ Updates from (35) and (17).
29 $\hat{\rho}, \hat{\eta}, \hat{\beta} \leftarrow$ Updates from (32), (33) and (34).
30 **end**
31 Update the messages $m^{\text{MF}}_{f_{D_i}\to x_i}(x_i)$ from (36).
32 Iterate message-passing in the BP subgraph.
33 Update the beliefs $q(x_i)$ from (37).
34 **end**

loop alternates between these two steps until the information bit estimates have not changed in 10 iterations or a maximum of 50 iterations is reached.

The scheduling of the channel estimation is inspired by [30]. The basic idea is to construct a representation of the CFR in (11) by sequential refinement of the estimated multipath components. One component is determined by the parameters (z_l, α_l, τ_l) for a particular index l. All multipath components are initialized in the inactivated state, i.e., $\hat{\boldsymbol{z}}$ is the zero vector.

The channel estimation procedure alternates between two stages: In the activation stage (at line 7) one of the inactive components is activated and its multipath delay and coefficient are calculated. The activation criterion (29) determines if the component should stay activated. In the second stage (starting at line 20), all active components are sequentially refined. Again, the criterion (29) determines if a component should be deactivated. The channel estimation procedure thus iteratively adds, updates and possibly removes components until the stopping criterion is fulfilled. The multipath delays are tracked via the scheme in Sec. IV-A2 in a way that resembles the operation of a rake receiver [17]. The approach presented here differs from that implemented in a rake receiver in that it provides an integral criterion for inclusion or exclusion of components (rake "fingers") via (29). The multipath delay of the newly activated component is found via a maximization over the grid $\tilde{\boldsymbol{\tau}}$. The grid should have a sufficiently fine resolution, such that the initial estimate of the delay is close to the maximizer in (26). We choose the distance between points in the grid as $(N\Delta_f)^{-1}/8$. As inner stopping criterion we use $|1/\hat{\beta}^{[t]} - 1/\hat{\beta}^{[t-1]}| < 10^{-3}/\hat{\beta}^{[t-1]}$, where t is the inner iteration number. The number of inner iterations is limited to 50.

During the first outer iteration the decoder has not been used yet and symbol beliefs $q(x_i)$ of the data subcarriers (indices $i \in \mathcal{D}$) are not available. During the first iteration the channel estimator therefore only uses the pilot subcarriers (indices $j \in \mathcal{P}$). To avoid any identifiability issue regarding the multipath delays (see Sec. V-C) during the pilot-only iteration, the multipath delays estimated in this iteration are restricted to the interval $[0, 1/(\Delta_f\Delta_P))$, where Δ_P is the pilot spacing.[12]

The active component prior variance is initialized to $\hat{\eta} = 1$ and the activation probability is initialized to $\hat{\rho} = 0.5$. We initialize the noise variance to $\hat{\beta} = \|\boldsymbol{y}\|^2/N \cdot 10^{-15/10}$ (i.e., assuming approximately 15 dB SNR). The activation probability and noise variance is kept fixed during the first 3 outer iterations, because these can only be accurately estimated when a reliable estimate of the channel is available.

D. Convergence Analysis and Computational Complexity

We now wish to analyze the convergence properties of Algorithm 1. First recognize that the algorithm alternates between updates in the MF and BP subgraphs of Fig. 3. To analyze convergence, we discuss under which conditions each of these sets of updates are guaranteed not to increase the RBFE. If all updates give a non-increasing RBFE it can be concluded that the algorithm converges, since the RBFE is bounded below.

We first discuss the updates in the MF subgraph, i.e., of belief functions $q(\alpha_l)$ $(l \in \mathcal{L})$ and point estimates $(\hat{\boldsymbol{z}}, \hat{\boldsymbol{\tau}}, \hat{\rho}, \hat{\eta}, \hat{\beta})$. During these updates the messages $m^{\text{BP}}_{f_{M_i}\to x_i}(x_i)$ are kept fixed. The joint update of $\hat{\tau}_l$ and $q(\alpha_l)$ gives a non-increasing RBFE as per Lemma 1. A similar conclusion can be drawn regarding the joint update of $\hat{z}_l$ and $q(\alpha_l)$. The individual update of $q(\alpha_l)$ is found via the method of Lagrange multipliers applied to the RBFE with normalization constraint $\int q(\alpha_l)\,\mathrm{d}\alpha_l = 1$. The second-order functional derivative of the RBFE $\frac{\delta^2 F_{\text{BP-MF}}}{\delta q^2(\alpha_l)} = \frac{1}{q(\alpha_l)}$ is a positive semi-definite function; it follows that the RBFE is convex in this argument. It can be concluded that the update of $q(\alpha_l)$ is the global minimizer of the RBFE and the objective is thus non-increasing. A similar conclusion can be drawn regarding the update of the channel parameters, cf. Eq. (30). All updates in the MF subgraph thus give non-increasing RBFE.

We now analyze the convergence in the BP subgraph, i.e., the updates of belief functions $q(x_i), q([\mathbf{c}^{(i)}]_q)$ and $q(u_k)$.

[12]We define the pilot spacing as $\Delta_P = D + 1$, where D is the number of data subcarriers between any two neighboring pilot subcarriers.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2868314, IEEE Transactions on Signal Processing

Parameter	Scenario A	Scenario B
Channel model	ITU-R M.2135 UMa NLOS [60]	IEEE 802.15.a Outdoor NLOS [61]
Number of subcarriers (N)	1024	1024
Modulation format of data subcarriers	256-QAM	256-QAM
Convolutional code polynomial	$(561, 753)_8$	$(561, 753)_8$
Subcarrier spacing (Δ_f)	25 kHz	250 kHz
Cyclic prefix duration (T_{CP})	5200 ns	800 ns
Number of equispaced pilots	172	256
Pilot spacing (Δ_P) (implied by the above)	6	4

TABLE I
SIMULATION PARAMETERS.

Considering the belief functions of variables in the MF subgraph as fixed and ignoring scaling and constant terms, the RBFE is equal to the Bethe free energy corresponding to the factorization (see [8, Appendix E])

$$\prod_{i\in\mathcal{D}} m^{\text{MF}}_{f_{D_i}\to x_i}(x_i)p(x_i|\boldsymbol{c}^{(i)})p(\boldsymbol{c}|\boldsymbol{u}) \prod_{k\in\mathcal{K}} p(u_k).$$

Further, all messages in the BP subgraph are equal to the messages obtained from BP applied to the above factorization. This means that we can analyze the behaviour of message-passing in the BP subgraph, by analyzing BP applied to the above factorization. If the factor graph does not contain any cycles it can be shown that BP globally minimizes the Bethe free energy [8], [51] (which in this case is equal to the Gibss free energy) and convergence of the complete BP-MF receiver algorithm is guaranteed. Recall that the factor $f_C(\boldsymbol{c}, \boldsymbol{u}) = p(\boldsymbol{c}|\boldsymbol{u})$ describes the channel code and may be replaced by a number of auxiliary variables and factors. The specific structure of the BP factor graph is thus determined by the channel code. In the special case of convolutional coding with binary or quadrature phase-shift keying (BPSK or QPSK) modulation, the BP graph does indeed become a tree-graph and convergence of Alg. 1 is guaranteed. If the modulation order is higher than QPSK, loops occur between f_{M_i} and f_C and convergence can thus not be guaranteed.

For other common channel codes, such as Turbo and LDPC codes, the subgraph represented by f_C contains loops. However, BP has empirically been shown to converge for decoding of many channel codes and it is a well known practice to use BP even though convergence cannot be guaranteed theoretically, see e.g. [55]–[58]. When BP does converge it has been shown to be to a (local) minimum of the Bethe free energy [62], which further explains why we do indeed obtain convergence of Alg. 1 in our numerical investigations. Conditions exist under which BP is guaranteed to converge in loopy graphs, e.g. [63], [64]. These are, however, not applicable to our situation.

We now turn our attention to the computational complexity of the channel estimator, i.e., the loop starting at line 5. The most demanding part of the channel estimation in terms of computational complexity is the calculation of $\hat{\boldsymbol{\mu}}_{\hat{\mathcal{A}}}$ via (35). We show in Appendix B that (under a conjecture) this update can be calculated in time $\mathcal{O}\Big(\min(\hat{L}^2N, \hat{L}N\sqrt{N})\Big)$, where $\hat{L}$ is the number of components currently included in the model.

The grid search in line 11 is recognized as the maximization of the periodogram, which can be calculated via a fast Fourier transform in time $\mathcal{O}(N \log N)$ when the grid is assumed to be of size $\mathcal{O}(N)$.

The loop starting at line 21 necessitates the calculation of $\boldsymbol{r}$ in (20). Direct computation has complexity $\mathcal{O}(\hat{L}N)$ for each of the $\hat{L}$ iterations in the loop. By updating $\boldsymbol{r}$ with each change to $\hat{\boldsymbol{\mu}}$, the direct evaluation can be avoided and the complexity of each iteration in the loop becomes $\mathcal{O}(N)$, which is the same as that of all other operations inside the loop. The overall complexity of the loop is thus $\mathcal{O}(\hat{L}N)$.

With these remarks, we see that the overall complexity per iteration of the channel estimator is $\mathcal{O}\Big(\min(\hat{L}^2N, \hat{L}N\sqrt{N})\Big)$.

V. NUMERICAL EVALUATION

In our numerical evaluation we consider an OFDM system as described in Sec. II. We use a random interleaver and a rate–1/2 non-systematic convolutional channel code, decoded by the loopy BP implementation from the Coded Modulation Library.[13] The pilot signals are chosen at random from a QPSK alphabet. The first and last subcarriers are designated as pilots. The other pilot subcarriers are located equispaced[14] with spacing Δ_P, i.e., the number of data subcarriers between two such neighbour pilot subcarriers is $\Delta_P - 1$. The SNR is defined based on the realization of the CFR as

$$\text{SNR} \triangleq \frac{\mathbb{E}\big[|x_i|^2\big]\,\|\boldsymbol{g}\|^2}{N\beta}, \tag{38}$$

where $\mathbb{E}\big[|x_i|^2\big]$ is calculated under the assumption that the symbols in the respective alphabets $\mathbb{A}_{\text{D}}$ and $\mathbb{A}_{\text{P}}$ are equiprobable.

We asses how the receivers behave in two different scenarios. The parameters considered in each scenario are listed in Table I. Scenario A uses the channel model put forward by ITU for the evaluation of IMT-Advanced radio interface technologies [60]. Specifically we use the model with the parameter setting for urban macro (UMa) environment with non line-of-sight (NLOS) conditions. The model generates impulse responses $h(\tau)$ typical of macro-cellular communication in an urban environment targeting continuous coverage for

[13] Available from http://iterativesolutions.com/Matlab.htm

[14] We have also conducted experiments with random pilot patterns (not shown), but have seen no significant benefit in doing so for the setup considered here.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2868314, IEEE Transactions on Signal Processing

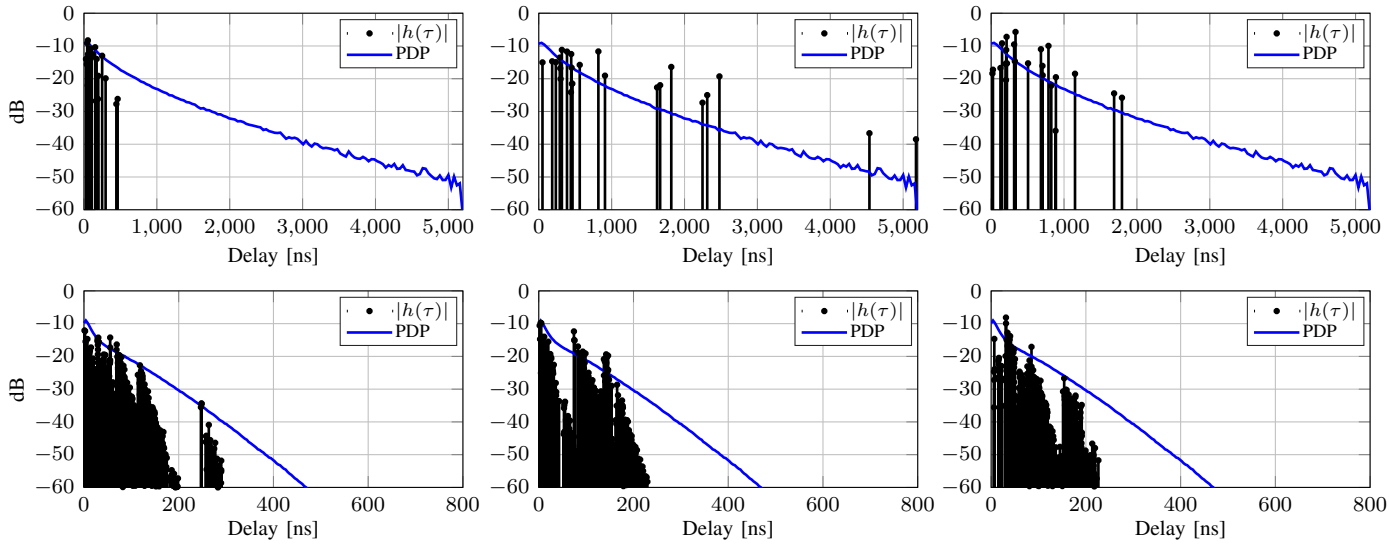

Fig. 4. Three sample realizations of $h(\tau)$ for Scenario A (top) and Scenario B (bottom). An estimate of the PDP is also shown, which is obtained by averaging the magnitude-squared impulse responses of $50,000$ channel realizations.

pedestrian up to fast vehicular users [60]. The channel model [60] is specified for use with up to 100 MHz bandwidth, while the system we are simulating uses 25.6 Mhz bandwidth. We are thus well within the specified bandwidth range.

Scenario B uses the standardized model proposed for the evaluation of IEEE 802.15.4a UWB technologies [61]. Specifically we use the model with the setting proposed for outdoor environments with NLOS conditions. The model generates impulse responses $h(\tau)$ typical of micro-cellular communication in a suburban-like environment with a rather small range [61]. Note that this model is also used in [11].

Since our signal model (6) is not valid for CIRs longer than the cyclic prefix duration T_{CP}, we drop realizations of the impulse response $h(\tau)$ with component delays larger than T_{CP}. Fig. 4 shows 3 impulse responses generated for each of scenarios A and B, along with an estimate of the PDP. An investigation of a few realizations has shown that in Scenario A most (but not all) pairs of neighbouring multipath components adhere to a separation by at least the reciprocal of the system bandwidth $1/(N\Delta_f)$. On the other hand, in Scenario B, there are many pairs of neighbouring multipath components that are not even separated by $10^{-2}/(N\Delta_f)$. In conclusion the impulse responses in Scenario A generally show a specular behaviour, while in Scenario B they show a dense behaviour.

We asses the performance of the considered receivers in terms of average coded bit error rate (BER) and normalized mean squared error (MSE) of the CFR, calculated as $\|\hat{\boldsymbol{g}} - \boldsymbol{g}\|^2/\|\boldsymbol{g}\|^2$. These averages are obtained from 500 Monte Carlo trials ($\approx 1.5\cdot 10^6$ information bits) for SNR < 20 dB, with one OFDM symbol transmitted in each trial. To get reliable BER estimates we use $3,000$ trials ($\approx 10^7$ information bits) for SNR $= 20$ dB and $15,000$ trials ($\approx 4.5\cdot 10^7$ information bits) for SNR > 20 dB. The OFDM symbols and channel realizations are generated i.i.d. according to the above.

A. Evaluation and Comparison with Other Algorithms

We evaluate our algorithm (Parametric BP-MF) and compare with the following reference algorithms:

Turbo-GAMP [11]: The algorithm employs a baud-spaced grid in the delay domain, i.e., the resolution of the grid is $T_s = (N\Delta_f)^{-1} \approx 39$ ns for Scenario A and $T_s \approx 3.9$ ns for Scenario B. For each channel tap a large-tap and small-tap variance is provided along with tap-state transition probabilities (see [11] for more details). These are estimated via the EM algorithm provided in [11] from $50,000$ channel realizations. Turbo-GAMP is provided with significant prior information on the CIR via these statistical values. We also provide Turbo-GAMP with the true noise variance, as [11] does not give a way to estimate this value.

LMMSE BP-MF [9]: The algorithm directly estimates the CFR $\boldsymbol{g}$ via the BP-MF framework. It is an iterative receiver with LMMSE channel estimation that requires prior knowledge of the noise variance and the covariance matrix $\mathbb{E}\{\boldsymbol{g}\boldsymbol{g}^{\text{H}}\}$. We provide the true noise variance to the receivers and show results using three different covariance matrices:

- A receiver using the covariance matrix calculated from the robust PDP described in [35], which assumes constant PDP within the interval $[0, T_{\text{CP}})$. This is known to be an appropriate choice when no statistical information about the channel is available at the receiver [35].
- A receiver using the true covariance matrix associated to the channel model. Due to the complex structure of the channel models, the true covariance matrix is not easy to obtain analytically. We therefore estimate it as the sample covariance matrix obtained from $50,000$ channel realizations. We identify this estimate with its true counterpart. The use of the true covariance matrix corresponds to knowing the true PDP (of which an estimate is shown in Fig. 4). We refer to this receiver as LMMSE BP-MF with known PDP.
- An oracle receiver that calculates the channel covariance matrix conditioned on (i.e., knowing) the true delays and powers of the multipath components. This oracle receiver is thus provided with significant side information. We refer to it as LMMSE BP-MF with multipath oracle.

These three choices of the covariance matrix progressively

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2868314, IEEE Transactions on Signal Processing

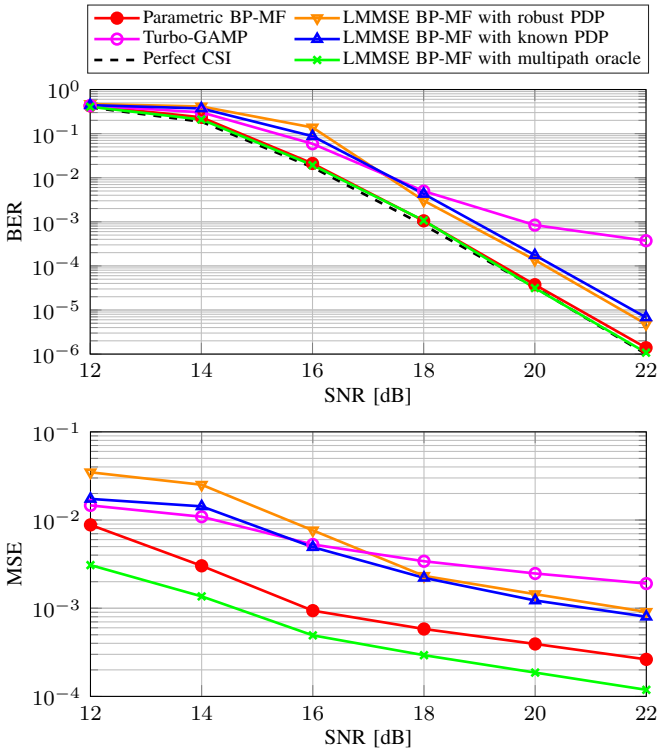

Fig. 5. BER (top) and MSE of CFR (bottom) vs. SNR in Scenario A.

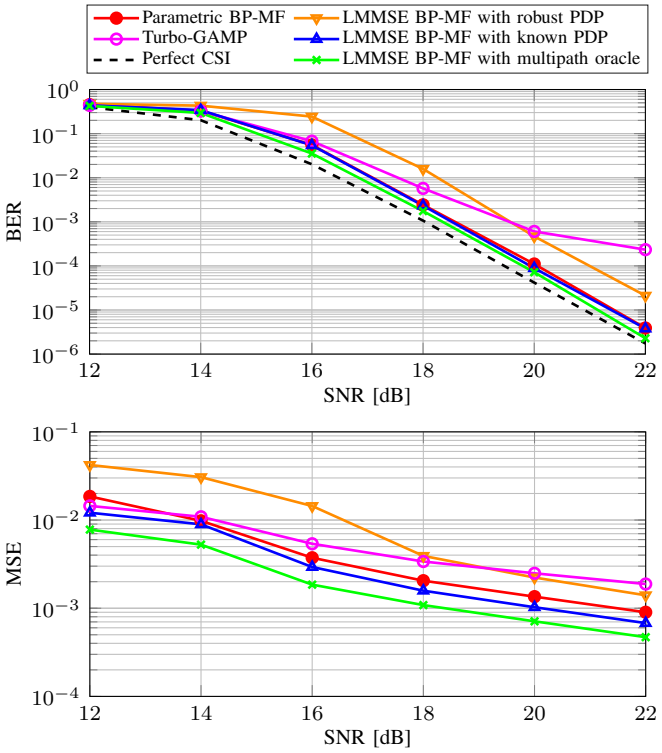

Fig. 6. BER (top) and MSE of CFR (bottom) vs. SNR in Scenario B.

assume stronger prior knowledge or side information to be available at the receiver.

Perfect CSI: This oracle receiver has perfect channel state information (CSI), i.e., it knows the true CFR $\boldsymbol{g}$, and thus provides a lower bound on the achievable BER. The Perfect CSI trace is only shown in the BER plots. To be specific, it is implemented by computing the messages $n_{x_i \to f_{M_i}}(x_i)$ for all $i \in \mathcal{D}$ (see (36)), followed by 5 iterations in the BP subgraph of Fig. 3.

B. Varying the Signal-to-Noise Ratio

Fig. 5 shows performance results for varying SNR in Scenario A. We first note that Parametric BP-MF performs very well in both BER and MSE. Its BER is remarkably close to that of the two oracle estimators (LMMSE BP-MF with multipath oracle and Perfect CSI), indicating that there is very little margin for improvement of the algorithm in this scenario. The robust and known PDP versions of LMMSE BP-MF show higher BER than Parametric BP-MF, corresponding to a decrease in SNR of about 1 dB. They show almost the same BER performance because the delay spread in Scenario A is relatively large (cf. Fig. 4) and the robust PDP assumption is therefore realistic. Turbo-GAMP does not perform well and shows a BER floor at high SNR. The reason is discussed below.

Fig. 6 shows the corresponding results for Scenario B. We here observe that Parametric BP-MF has a BER loss compared to the Perfect CSI trace corresponding to about 0.5 dB SNR difference. Parametric BP-MF is among the best performing algorithms, even though the impulse responses generated in Scenario B are dense and thus composed of a very large number of multipath components that the algorithm cannot resolve individually (cf. Fig. 4). Instead, the algorithm estimates a virtual CIR with significantly fewer components that approximates the true CIR within the system bandwidth. We have observed that the estimated virtual CIR approximately "recovers the support" of the true CIR, in the sense that an estimated multipath component is located wherever the CIR contains significant power. Parametric BP-MF has a BER and MSE performance equivalent to that of LMMSE BP-MF with both known PDP and multipath oracle. We stress that Parametric BP-MF achieves this performance without using prior knowledge of the channel.

In Scenario B we observe a significant difference between the LMMSE BP-MF algorithms with known and robust PDP. To explain this difference observe in Fig. 4 that most of the mass of the PDP is located at small delays. This significantly deviates from the evenly distributed mass on $[0, T_{\mathrm{CP}})$ that underlies the robust PDP assumption.

In both Fig. 5 and Fig. 6 an error floor is observed for Turbo-GAMP at high SNR.[15] We conjecture that this error floor is caused by the restriction of the delays to the baud-spaced grid. If the delays are generated to be located on that grid, the performance of Turbo-GAMP is very close to that of the Perfect CSI trace (not shown here). We have also conducted experiments with random pilot patterns (not shown) as used in [11] (where Turbo-GAMP is introduced) but did not see an

[15] In [11], where Turbo-GAMP is introduced, such an error floor is not observed even though the setup in the numerical investigation is almost identical to that in Scenario B. The reason is an error in the signal model in [11] that invalidates the numerical results obtained in that paper. Specifically the error occurs when the "uniformly sampled channel taps" are defined as rate $1/T$ samples of the compound CIR $x(\tau) \triangleq (g_r * h * g_t)(\tau)$ (notation as in [11]). However, since $(g_r * g_t)(\tau)$ is a raised-cosine filter with design parameter 0.5, $x(\tau)$ has bandwidth $1.5/T$, leading to aliasing in the sampling operation.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2868314, IEEE Transactions on Signal Processing

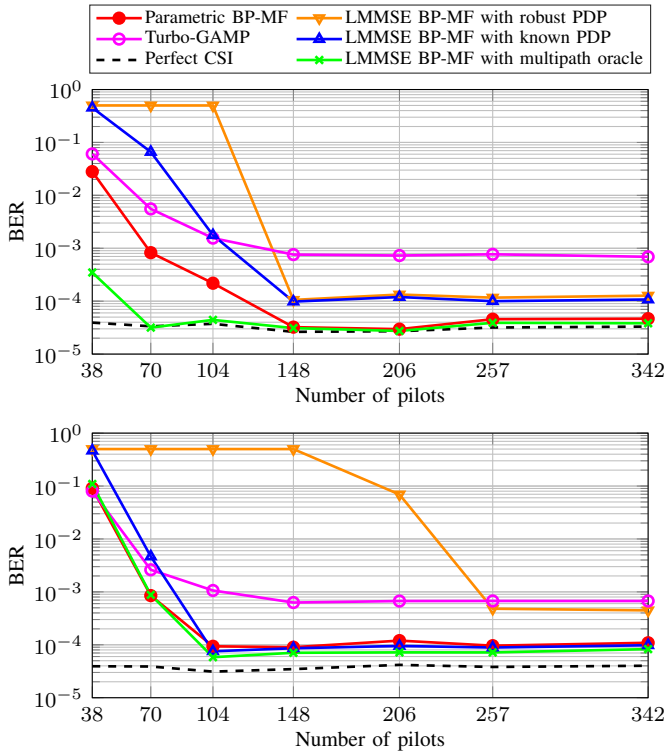

Fig. 7. BER vs. number of pilot subcarriers in Scenario A (top) and Scenario B (bottom) at 20 dB SNR.

improvement of Turbo-GAMP in that case. We note that such error floors in BER and MSE have previously been observed for other grid-based sparse channel estimation algorithms, see for example [21], [43]. In conclusion, the grid-based approximation is of insufficient accuracy for communication with large modulation order in the high-SNR regime.

C. Varying the Number of Pilots

We now investigate if our receiver design improves the trade-off between the number of pilots and estimator performance. Fig. 7 shows the BER performance for varying number of pilot subcarriers.

The first observation is that LMMSE BP-MF with robust PDP shows a point at which the BER performance quickly transitions between high (50 %) BER and low ($< 10^{-3}$) BER. Under the robust PDP assumption, the channel coherence bandwidth is approximately $1/T_{\mathrm{CP}}$. As a rule of thumb there should be at least one pilot subcarrier per coherence interval, which gives the criterion $P > N\Delta_f T_{\mathrm{CP}}$, where P is the needed number of pilot subcarriers. For Scenario A we have $N\Delta_f T_{\mathrm{CP}} \approx 133$ and for Scenario B we have $N\Delta_f T_{\mathrm{CP}} \approx 205$, which exactly are the respective numbers of pilots at which LMMSE BP-MF with robust PDP transitions between low and high BER.

All algorithms except LMMSE BP-MF with robust PDP can operate significantly below the above-mentioned limit. Due to the iterative processing, the number of pilots can be decreased significantly without incurring an increase in BER.

VI. Conclusions

In this paper we proposed an iterative OFDM receiver that employs sparsity-based parametric channel estimation. The iterative receiver is derived using the BP-MF framework for approximate Bayesian inference. Unlike state-of-the-art sparse channel estimators, our scheme does not restrict multipath delays of the estimated channel impulse response to a grid. As a result it can truly exploit parsimony of the channel impulse response, without resorting to approximate sparsity (as in [11], [23], [24]).

We have presented a numerical evaluation that compares our algorithm with state-of-the-art methods, i.e., Turbo-GAMP [11] and LMMSE BP-MF [9]. This study demonstrated that restricting the multipath delays to a baud-spaced grid (e.g., as in Turbo-GAMP) is not a viable approach because the resulting equivalent vector of channel taps is only approximately sparse.

The numerical evaluation also shows that our proposed scheme can effectively exploit the structure of wireless channel impulse responses. We have showed numerically that parametric channel estimation works well with both specular and dense channels. Our analysis of the channel covariance matrix in Sec. II-B shows that for dense channels a virtual channel impulse response can be estimated, with a number of virtual components given by the (effective) rank of the channel covariance matrix. The corresponding virtual frequency response approximates the actual channel frequency response well within the system bandwidth.

Appendix A
The Region-Based Free Energy Approximation

At the heart of the derivation of our algorithm lies the RBFE as defined by [8, Eq. (17)], [51]. In this paper we use the RBFE of the probability distribution corresponding to the factor graph depicted in Fig. 3. For convenience, we give here the complete expression of the RBFE:

$$F_{\text{BP-MF}} = F_{\text{BP}} + F_{\text{MF}} \tag{39}$$

with

$$\begin{aligned}
F_{\text{BP}} = &\sum_{k\in\mathcal{K}} \sum_{u_k\in\{0,1\}} b_{u_k}(u_k) \ln \frac{b_{u_k}(u_k)}{p(u_k)} \\
&+ \sum_{i\in\mathcal{D}} \sum_{\substack{x_i\in\mathbb{A}_{\mathrm{D}} \\ \boldsymbol{c}^{(i)}\in\{0,1\}^Q}} b_{M_i}(x_i,\boldsymbol{c}^{(i)}) \ln \frac{b_{M_i}(x_i,\boldsymbol{c}^{(i)})}{p(x_i|\boldsymbol{c}^{(i)})} \\
&+ \sum_{\substack{\boldsymbol{c}\in\{0,1\}^{K/R} \\ \boldsymbol{u}\in\{0,1\}^K}} b_C(\boldsymbol{c},\boldsymbol{u}) \ln \frac{b_C(\boldsymbol{c},\boldsymbol{u})}{p(\boldsymbol{c}|\boldsymbol{u})} - \sum_{k\in\mathcal{K}} \sum_{u_k\in\{0,1\}} q(u_k) \ln q(u_k) \\
&- \sum_{i\in\mathcal{D}} \sum_{m\in\{1,\ldots,Q\}} \sum_{[\boldsymbol{c}^{(i)}]_m\in\{0,1\}} q([\boldsymbol{c}^{(i)}]_m) \ln q([\boldsymbol{c}^{(i)}]_m),
\end{aligned}$$

$$\begin{aligned}
F_{\text{MF}} = &\sum_{l\in\mathcal{L}} \int q(\alpha_l) \ln q(\alpha_l)\, \mathrm{d}\alpha_l \\
&- \left\langle \ln p(\boldsymbol{y}|\boldsymbol{x}_{\mathcal{D}},\boldsymbol{\alpha},\hat{\boldsymbol{\tau}};\hat{\beta}) \right\rangle_{\boldsymbol{x}_{\mathcal{D}},\boldsymbol{\alpha}} \\
&- \sum_{l\in\mathcal{L}} \left\langle \ln p(\alpha_l|\hat{z}_l;\hat{\eta}) p(\hat{z}_l;\hat{\rho}) p(\hat{\tau}_l) \right\rangle_{\alpha_l},
\end{aligned}$$

where $b_C(\boldsymbol{c},\boldsymbol{u})$, $b_{M_i}(x_i,\boldsymbol{c}^{(i)})$ for $i\in\mathcal{D}$ and $b_{u_k}(u_k)$ for $k\in\mathcal{K}$ are factor beliefs. With abuse of notation we let $q(\cdot)$ denote variable beliefs and $\langle\cdot\rangle_a$ denote expectation with respect to the belief density $q(a)$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2868314, IEEE Transactions on Signal Processing

APPENDIX B
EFFICIENT CALCULATION OF $\hat{\boldsymbol{\mu}}_{\hat{\mathcal{A}}}$ WHEN $\hat{L}$ IS LARGE

In this appendix we present a computationally efficient method for evaluating $\hat{\boldsymbol{\mu}}_{\hat{\mathcal{A}}}$ as defined by (35). We here present an iterative approach but note that an alternative (non-iterative) fast method can most likely be obtained by extending the approach of [65].

Direct evaluation and inversion of $\mathbf{Q}$ has time complexity $\mathcal{O}(\hat{L}^2 N)$, where $\hat{L} \triangleq |\hat{\mathcal{A}}|$. The iterative method presented here has complexity $\mathcal{O}(\hat{L}N\sqrt{N})$ provided Conjecture 1 (below) holds. It is thus beneficial to use it when $\hat{L}$ grows faster than $\sqrt{N}$.

We first use the Woodbury matrix identity to write $\hat{\boldsymbol{\mu}}$ as

$$\hat{\boldsymbol{\mu}} = \hat{\beta}^{-1}\hat{\eta}\left(\mathbf{I} - \hat{\beta}^{-1}\hat{\eta}\boldsymbol{\Psi}^{\mathrm{H}}(\hat{\boldsymbol{\tau}}_{\hat{\mathcal{A}}})\mathbf{C}^{-1}\boldsymbol{\Psi}(\hat{\boldsymbol{\tau}}_{\hat{\mathcal{A}}})\right)\boldsymbol{\Psi}^{\mathrm{H}}(\hat{\boldsymbol{\tau}}_{\hat{\mathcal{A}}})\left\langle \mathbf{X}\right\rangle_{\boldsymbol{x}_{\mathcal{D}}}^{\mathrm{H}}\boldsymbol{y},$$

where

$$\mathbf{C} = \left\langle \mathbf{X}^{\mathrm{H}}\mathbf{X}\right\rangle_{\boldsymbol{x}_{\mathcal{D}}}^{-1} + \hat{\beta}^{-1}\hat{\eta}\boldsymbol{\Psi}(\hat{\boldsymbol{\tau}}_{\hat{\mathcal{A}}})\boldsymbol{\Psi}^{\mathrm{H}}(\hat{\boldsymbol{\tau}}_{\hat{\mathcal{A}}}).$$

We immediately recognize that the computationally dominating part is to solve a system of N linear equations of the form $\mathbf{C}\boldsymbol{z} = \boldsymbol{a}$. Since $\mathbf{C}$ is Hermitian and positive-definite, we can solve this system via the conjugate-gradient (CG) method (Alg. 2.1 in [66]), which is an iterative method for solving systems of linear equations. In the following we show that the number of iterations of the CG method is $\mathcal{O}(\sqrt{N})$.

We first need a conjecture on the eigenvalues of the (Hermitian-Toeplitz) matrix $\mathbf{T} = \hat{\beta}^{-1}\hat{\eta}\boldsymbol{\Psi}(\hat{\boldsymbol{\tau}}_{\hat{\mathcal{A}}})\boldsymbol{\Psi}^{\mathrm{H}}(\hat{\boldsymbol{\tau}}_{\hat{\mathcal{A}}})$.

Conjecture 1: There exists an upper bound on the largest eigenvalue of $\mathbf{T}$ that grows linearly with N, i.e.,

$$\lambda_{\max}(\mathbf{T}) = \mathcal{O}(N).$$

To justify this conjecture we refer to Fig. 8, where the largest eigenvalue is shown for varying N.

We also need a number of lemmas.

Lemma 2: There exists constants $c_1 > 0$ and $c_2 < \infty$, such that $c_1 \leq \left\langle |x_i|^2\right\rangle_{\boldsymbol{x}_{\mathcal{D}}} \leq c_2$ for all $i \in \mathcal{D} \cup \mathcal{P}$.

Proof: Observe that the data and pilot modulation symbol alphabets $\mathbb{A}_{\mathrm{D}}$ and $\mathbb{A}_{\mathrm{P}}$ only contain finite, non-zero values. We can thus take $c_1 = \min_{x\in\mathbb{A}_{\mathrm{P}}\cup\mathbb{A}_{\mathrm{D}}}|x|^2$ and $c_2 = \max_{x\in\mathbb{A}_{\mathrm{P}}\cup\mathbb{A}_{\mathrm{D}}}|x|^2$ to complete the proof. ∎

Lemma 3: Assume that Conjecture 1 holds. The largest and smallest eigenvalues of $\mathbf{C}$ obey

$$\lambda_{\max}(\mathbf{C}) = \mathcal{O}(N), \qquad \lambda_{\min}(\mathbf{C}) \geq c_1^{-1}.$$

Proof: By the Weyl inequality for Hermitian matrices $\mathbf{C}$, $\mathbf{T}$ and $\left\langle \mathbf{X}^{\mathrm{H}}\mathbf{X}\right\rangle_{\boldsymbol{x}_{\mathcal{D}}}^{-1}$ we have

$$\lambda_{\max}(\mathbf{C}) \leq \lambda_{\max}\left(\left\langle \mathbf{X}^{\mathrm{H}}\mathbf{X}\right\rangle_{\boldsymbol{x}_{\mathcal{D}}}^{-1}\right) + \lambda_{\max}(\mathbf{T}).$$

The first inequality follows directly from Conjecture 1 and Lemma 2.

Similarly by the dual Weyl inequality

$$\lambda_{\min}(\mathbf{C}) \geq \lambda_{\min}\left(\left\langle \mathbf{X}^{\mathrm{H}}\mathbf{X}\right\rangle_{\boldsymbol{x}_{\mathcal{D}}}^{-1}\right) + \lambda_{\min}(\mathbf{T}).$$

Since $\hat{L} < N$, the matrix $\mathbf{T}$ is singular and $\lambda_{\min}(\mathbf{T}) = 0$. The second inequality now follows from Lemma 2. ∎

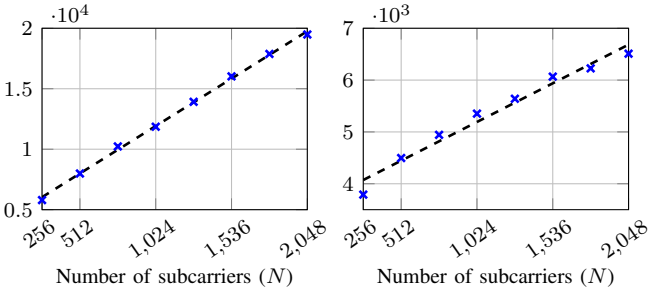

Fig. 8. Average of the largest eigenvalue of the matrix $\mathbf{T}$ encountered during one execution of Parametric BP-MF for Scenario A (left) and B (right). Average obtained from 1000 Monte Carlo trials. Both plots were generated at 20 dB SNR. A dashed line depicts the least-squares linear fit.

By Theorem 2.2 in [66] the number of iterations required by the CG method to achieve a desired accuracy in the solution of $\boldsymbol{a} = \mathbf{C}\boldsymbol{z}$ is $\mathcal{O}\left(\sqrt{\frac{\lambda_{\max}(\mathbf{C})}{\lambda_{\min}(\mathbf{C})}}\right)$. By Lemma 3 the number of iterations is thus $\mathcal{O}(\sqrt{N})$. Each iteration has time complexity $\mathcal{O}(\hat{L}N)$ and the overall complexity of solving (35) via this method is therefore $\mathcal{O}(\hat{L}N\sqrt{N})$.

REFERENCES

[1] *Evolved Universal Terrestrial Radio Access (E-UTRA), Physical channels and modulation*, 3GPP Std. 36.211, Rev. 12.8.0, Dec. 2015.

[2] C. Douillard, M. Jézéquel, C. Berrou, A. Picart, P. Didier, and A. Glavieux, "Iterative correction of intersymbol interference: Turbo-equalization," *European Trans. Telecommun.*, vol. 6, pp. 507–511, 1995.

[3] M. Tüchler and A. C. Singer, "Turbo equalization: An overview," *IEEE Trans. Inf. Theory*, vol. 57, pp. 920–952, Feb. 2011.

[4] S. Park, Y. G. Kim, and C. G. Kang, "Iterative receiver for joint detection and channel estimation in OFDM systems under mobile radio channels," *IEEE Trans. Veh. Technol.*, vol. 53, pp. 450–460, Mar. 2004.

[5] A. P. Worthen and W. E. Stark, "Unified design of iterative receivers using factor graphs," *IEEE Trans. Inf. Theory*, vol. 47, pp. 843–849, Feb. 2001.

[6] S. Wu, L. Kuang, Z. Ni, J. Lu, D. David Huang, and Q. Guo, "Expectation propagation approach to joint channel estimation and decoding for OFDM systems," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process.*, May 2014, pp. 1941–1945.

[7] Y. Zhu, D. Guo, and M. L. Honig, "A message-passing approach for joint channel estimation, interference mitigation, and decoding," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 6008–6018, Dec. 2009.

[8] E. Riegler, G. E. Kirkelund, C. N. Manchón, M. Badiu, and B. H. Fleury, "Merging belief propagation and the mean field approximation: A free energy approach," *IEEE Trans. Inf. Theory*, vol. 59, pp. 588–602, Jan. 2013.

[9] M.-A. Badiu, G. E. Kirkelund, C. N. Manchón, E. Riegler, and B. H. Fleury, "Message-passing algorithms for channel estimation and decoding using approximate inference," in *Proc. IEEE Int. Symp. Inform. Theory*, Jul. 2012, pp. 2376–2380.

[10] P. Schniter, "Belief-propagation-based joint channel estimation and decoding for spectrally efficient communication over unknown sparse channels," *Physical Communication*, vol. 5, pp. 91–101, Jun. 2012.

[11] ——, "A message-passing receiver for BICM-OFDM over unknown clustered-sparse channels," *IEEE J. Select. Topics in Signal Process.*, vol. 5, pp. 1462–1474, Dec. 2011.

[12] L. Marple, "Resolution of conventional Fourier, autoregressive, and special ARMA methods of spectrum analysis," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 2, May 1977, pp. 74–77.

[13] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Commun. Pure and Appl. Math.*, vol. 67, pp. 906–956, 2014.

[14] M. Feder and J. A. Catipovic, "Algorithms for joint channel estimation and data recovery-application to equalization in underwater communications," *IEEE J. Ocean. Eng.*, vol. 16, pp. 42–55, Jan. 1991.

[15] V. Lottici, A. D'Andrea, and U. Mengali, "Channel estimation for ultra-wideband communications," *IEEE J. Sel. Areas Commun.*, vol. 20, pp. 1638–1645, Dec. 2002.

[16] I. Maravic, J. Kusuma, and M. Vetterli, "Low-sampling rate UWB channel characterization and synchronization," *J. of Commun. and Networks*, vol. 5, pp. 319–327, Dec. 2003.

[17] A. F. Molisch, A. Mammela, and D. P. Taylor, Eds., *Wideband Wireless Digital Communication*. Prentice Hall, 2000.

[18] W. Bajwa, A. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, pp. 1058–1076, Jun. 2010.

[19] J. Paredes, G. Arce, and Z. Wang, "Ultra-wideband compressed sensing: Channel estimation," *IEEE J. Sel. Topics Signal Process.*, vol. 1, pp. 383–395, Oct. 2007.

[20] C. Berger, S. Zhou, J. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing," *IEEE Trans. Signal Process.*, vol. 58, pp. 1708 – 1721, Mar. 2010.

[21] G. Taubock, F. Hlawatsch, D. Eiwen, and H. Rauhut, "Compressive estimation of doubly selective channels in multicarrier syst.: Leakage effects and sparsity-enhancing process." *IEEE J. Sel. Topics Signal Process.*, vol. 4, pp. 255–271, Apr. 2010.

[22] N. L. Pedersen, C. N. Manchón, D. Shutin, and B. H. Fleury, "Application of Bayesian hierarchical prior modeling to sparse channel estimation," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2012, pp. 3487–3492.

[23] R. Prasad, C. R. Murthy, and B. D. Rao, "Joint approximately sparse channel estimation and data detection in OFDM systems using sparse Bayesian learning," *IEEE Trans. Signal Process.*, vol. 62, pp. 3591–3603, Jun. 2014.

[24] ——, "Joint channel estimation and data detection in MIMO-OFDM systems: A sparse Bayesian learning approach," *IEEE Trans. Signal Process.*, vol. 63, pp. 5369–5382, Jun. 2015.

[25] J.-J. Van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Börjesson, "On channel estimation in OFDM systems," in *Proc. IEEE 45th Veh. Technology Conf.*, vol. 2, Jul. 1995, pp. 815–819.

[26] Y. Chi, L. Scharf, A. Pezeshki, and A. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Trans. Signal Process.*, vol. 59, pp. 2182–2195, May 2011.

[27] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Trans. Signal Process.*, vol. 61, pp. 5987–5999, Dec. 2013.

[28] S. Pejoski and V. Kafedziski, "Estimation of sparse time dispersive channels in pilot aided OFDM using atomic norm," *IEEE Wireless Commun. Lett.*, vol. 4, pp. 397–400, Apr. 2015.

[29] Y. Barbotin and M. Vetterli, "Fast and robust parametric estimation of jointly sparse channels," *IEEE J. on Emerging and Select. Topics in Circuits and Syst.*, vol. 2, pp. 402–412, Sep. 2012.

[30] T. L. Hansen, M. A. Badiu, B. H. Fleury, and B. D. Rao, "A sparse Bayesian learning algorithm with dictionary parameter estimation," in *Proc. IEEE 8th Sensor Array and Multichannel Signal Process. Workshop*, Jun. 2014, pp. 385–388.

[31] M.-A. Badiu, T. L. Hansen, and B. H. Fleury, "Variational Bayesian inference of line spectra," *IEEE Trans. Signal Process.*, vol. 65, pp. 2247–2261, May 2017.

[32] D. Shutin, W. Wang, and T. Jost, "Incremental sparse Bayesian learning for parameter estimation of superimposed signals," in *Proc. 10th Int. Conf. Sampling Theory and Applicat.*, Jul. 2013.

[33] L. Hu, Z. Shi, J. Zhou, and Q. Fu, "Compressed sensing of complex sinusoids: An approach based on dictionary refinement," *IEEE Trans. Signal Process.*, vol. 60, pp. 3809–3822, Jul. 2012.

[34] J. J. Kormylo and J. Mendel, "Maximum likelihood detection and estimation of Bernoulli-Gaussian processes," *IEEE Trans. Inf. Theory*, vol. 28, pp. 482–488, May 1982.

[35] O. Edfors, M. Sandell, J.-J. Van de Beek, S. K. Wilson, and P. O. Börjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Commun.*, vol. 46, pp. 931–939, Jul. 1998.

[36] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes," in *Proc. IEEE Int. Conf. Commun.*, vol. 2, May 1993, pp. 1064–1070.

[37] R. Gallager, "Low-density parity-check codes," *Inform. Theory, IRE Trans.*, vol. 8, pp. 21–28, January 1962.

[38] M. Luise, R. Reggiannini, and G. M. Vitetta, "Blind equalization/detection for OFDM signals over frequency-selective channels," *IEEE J. Sel. Areas Commun.*, vol. 16, pp. 1568–1578, Oct. 1998.

[39] P. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun. Syst.*, vol. 11, pp. 360–393, Dec. 1963.

[40] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*, 1st ed. Prentice Hall, 2005.

[41] C. Carathéodory and L. Fejér, "Über den zusammenhang der extremen von harmonischen funktionen mit ihren koeffizienten und über den picard-landau'schen satz," *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, vol. 32, pp. 218–239, Jan. 1911.

[42] H. Sari, G. Karam, and I. Jeanclaude, "Transmission techniques for digital terrestrial TV broadcasting," *IEEE Commun. Mag.*, vol. 33, pp. 100–109, Feb. 1995.

[43] O.-E. Barbu, N. L. Pedersen, C. N. Manchón, G. Monghal, C. Rom, and B. H. Fleury, "Sparse channel estimation including the impact of the transceiver filters with application to OFDM," in *Proc. 15th Int. Workshop on Signal Process. Advances in Wireless Commun.*, Jun. 2014, pp. 424–428.

[44] R. J.-M. Cramer, R. A. Scholtz, and M. Z. Win, "Evaluation of an ultra-wide-band propagation channel," *IEEE Trans. Antennas Propag.*, vol. 50, pp. 561–570, May 2002.

[45] A. F. Molisch, "Ultrawideband propagation channels – theory, measurement, and modelling," *IEEE Trans. Veh. Technol.*, vol. 54, pp. 1528–1545, Oct. 2004.

[46] M. Stojanovic and J. Preisig, "Underwater acoustic communication channels: Propagation models and statistical characterization," *IEEE Commun. Mag.*, vol. 47, pp. 84–89, Jan. 2009.

[47] C. R. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of compressive sensing to sparse channel estimation," *IEEE Commun. Mag.*, vol. 48, pp. 164–174, Nov. 2010.

[48] G. L. Turin, "Communication through noisy, random-multipath channels," Ph.D. dissertation, Massachusetts Institute of Technology, 1956.

[49] D. Slepian, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty — V: The discrete case," *Bell Labs Technical Journal*, vol. 57, pp. 1371–1430, May 1978.

[50] T. L. Hansen, "Sparsity-based algorithms for line spectral estimation," Ph.D. dissertation, Aalborg University, Denmark, Mar. 2018.

[51] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms," *IEEE Trans. Inf. Theory*, vol. 51, pp. 2282–2312, 2005.

[52] J. M. Winn and C. M. Bishop, "Variational message passing," *J. Mach. Learning Research*, vol. 6, pp. 661–694, Apr. 2005.

[53] M. J. Beal, "Variational algorithms for approximate bayesian inference," Ph.D. dissertation, University College London, 2003.

[54] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.

[55] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, pp. 498–519, Feb. 2001.

[56] R. McEliece, D. MacKay, and J.-F. Cheng, "Turbo decoding as an instance of Pearl's "belief propagation" algorithm," *IEEE J. Sel. Areas Commun.*, vol. 16, pp. 140–152, Feb 1998.

[57] F. R. Kschischang and B. J. Frey, "Iterative decoding of compound codes by probability propagation in graphical models," *IEEE J. Sel. Areas Commun.*, vol. 16, pp. 219–230, Feb. 1998.

[58] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. 20, pp. 284–287, Mar 1974.

[59] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed. The Johns Hopkins Univ. Press, 1996.

[60] *Guidelines for evaluation of radio interface technologies for IMT-Advanced*, ITU-R Std. M.2135-1, Dec. 2008.

[61] A. F. Molisch, D. Cassioli, C.-C. Chong, S. Emami, A. Fort, B. Kannan, J. Karedal, J. Kunisch, H. G. Schantz, K. Siwiak *et al.*, "A comprehensive standardized model for ultrawideband propagation channels," *IEEE Trans. Antennas Propag.*, vol. 54, pp. 3151–3166, Nov. 2006.

[62] T. Heskes, "Stable fixed points of loopy belief propagation are local minima of the Bethe free energy," in *Advances in neural information processing systems*, 2002, pp. 343–350.

[63] J. M. Mooij and H. J. Kappen, "Sufficient conditions for convergence of the sum–product algorithm," *IEEE Trans. Inf. Theory*, vol. 53, pp. 4422–4437, Dec. 2007.

[64] Q. Su and Y.-C. Wu, "On convergence conditions of Gaussian belief propagation," *IEEE Trans. Signal Process.*, vol. 63, pp. 1144–1155, May 2015.

[65] T. L. Hansen, B. H. Fleury, and B. D. Rao, "Superfast line spectral estimation," *IEEE Trans. Signal Process.*, vol. 66, pp. 2128–2143, May 2018.

[66] M. K. Ng, *Iterative Methods for Toeplitz Systems*. Oxford University Press, 2004.