Physical Communication ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at SciVerse ScienceDirect

Physical Communication

journal homepage: www.elsevier.com/locate/phycom

Full length article

Belief-propagation-based joint channel estimation and decoding for spectrally efficient communication over unknown sparse channels

Philip Schniter*

Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH, United States

ARTICLE INFO

Article history:
Received 17 December 2010
Received in revised form 27 May 2011
Accepted 10 September 2011
Available online xxxx

Keywords:
Belief propagation
Message passing
Compressive sensing
Compressed sensing
Sparse channels
OFDM

ABSTRACT

We consider spectrally-efficient communication over a Rayleigh N-block-fading channel with a K-sparse L-length discrete-time impulse response (for $0 < K < L < N$), where neither the transmitter nor the receiver know the channel's coefficients nor its support. Since the high-SNR ergodic capacity of this channel has been shown to obey $C(\mathsf{SNR}) = (1-K/N)\log_2(\mathsf{SNR}) + \mathcal{O}(1)$, any pilot-aided scheme that sacrifices more than K dimensions per fading block to pilots will be spectrally inefficient. This causes concern about the conventional "compressed channel sensing" approach, which uses $\mathcal{O}(K\,\mathrm{polylog}(L))$ pilots. In this paper, we demonstrate that practical spectrally-efficient communication is indeed possible. For this, we propose a novel belief-propagation-based reception scheme to use with a standard bit-interleaved coded orthogonal frequency division multiplexing (OFDM) transmitter. In particular, we leverage the "relaxed belief propagation" methodology, which allows us to perform joint sparse-channel estimation and data decoding with only $\mathcal{O}(LN)$ complexity. Empirical results show that our receiver achieves the desired capacity pre-log factor of $1-K/N$ and performs near genie-aided bounds at both low and high SNR.

1. Introduction

Our goal is to communicate, in a spectrally efficient manner, over a Rayleigh N-block-fading channel with a K-sparse discrete-time impulse response of length L (where $0 < K < L < N$), under the realistic assumption that neither the transmitter nor the receiver knows the channel's coefficients nor its support. It has been recently shown [1] that the ergodic capacity of this noncoherent sparse channel obeys

$$C_{\text{sparse}}(\mathsf{SNR}) = \frac{N-K}{N}\log_2(\mathsf{SNR}) + \mathcal{O}(1) \tag{1}$$

as the signal-to-noise ratio (SNR) grows large. For comparison, the high-SNR ergodic noncoherent capacity of the Rayleigh N-block-fading L-length *non*-sparse channel obeys [2]

$$C_{\text{non-sparse}}(\mathsf{SNR}) = \frac{N-L}{N}\log_2(\mathsf{SNR}) + \mathcal{O}(1), \tag{2}$$

which exhibits a lower pre-log factor than (1). Thus, information theory confirms that channel sparsity can indeed be exploited to increase spectral efficiency, at least for high SNR. In particular, it establishes that, in the high-SNR regime, the signaling scheme does not need to sacrifice more than K degrees-of-freedom per fading-block to mitigate the effects of not knowing the K non-zero channel coefficients nor their locations.

Among the many strategies that exist for communication over unknown channels, pilot-aided transmission (PAT) [3] has emerged as one of the most effective. For example, it is known [2] that, for the Rayleigh N-block-fading L-length *non*-sparse channel, PAT achieves rates in

* Correspondence to: Department of ECE, 2015 Neil Ave., Columbus OH 43210, United States. Tel.: +1 614 247 6488; fax: +1 614 292 7596.
E-mail address: schniter@ece.osu.edu.

1874-4907/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.phycom.2011.09.004

accordance with the capacity expression (2).[1] It is then not surprising that the vast majority of techniques that have recently been proposed for communication over *sparse* channels are also based on PAT (see, e.g., the extensive bibliography in [4]). Broadly speaking, these techniques propose to exploit channel sparsity in order to reduce the number of pilots used for accurate channel estimation, with the end goal of increasing spectral efficiency. Typically, these schemes take a *decoupled* approach to reception: a sparse-channel estimate is calculated from pilot observations using a practical compressed sensing algorithm like LASSO [5,6], and the channel estimate is subsequently used for data decoding. Hereafter, we shall refer to this decoupled approach as "*compressed channel sensing*" (CCS), after [4]. When $\mathcal{O}(K\text{polylog}(L))$ pilots[2] are used for CCS, the theory of *compressed sensing* guarantees that – with high probability – the resulting channel estimates will be accurate, e.g., their squared-error will decrease in proportion to the received noise variance [4].

While the use of $\mathcal{O}(K\text{polylog}(L))$ pilots may be an improvement over L pilots required when channel sparsity is not taken into account, the capacity expression (1) implies that any PAT scheme sacrificing more than K degrees of freedom (per fading block) to pilots will be spectrally inefficient in the high-SNR regime. Thus, any scheme based on CCS, which uses $\mathcal{O}(K\text{polylog}(L)) > K$ pilots, will fall short of maximizing spectral efficiency. One may then wonder whether there exists a practical[3] communication scheme that achieves the capacity prelog factor in (1).

In this paper, we propose a novel approach to communication over sparse channels that (empirically) achieves rates in accordance with the sparse-channel capacity expression (1). For transmission, we use a conventional scheme, based on bit-interleaved coded modulation (BICM) with orthogonal frequency division multiplexing (OFDM) and a few carefully placed training bits. For reception, we deviate from the CCS approach and perform sparse-channel estimation and data decoding *jointly*. To accomplish this latter task in a practical manner, we take an approach suggested by *belief-propagation* (BP) [7], leveraging recent advances in "relaxed BP" [8,9] and in BP-based soft-input/soft-output (SISO) decoding [10]. The scheme that we propose has very low computation complexity: only $\mathcal{O}(NL)$ multiplies per fading block are required. Thus, we are able to handle long channels, many subcarriers, and large QAM constellations (which are in fact necessary to achieve high spectral efficiencies). Our simulations, for example, use $N = 1021$ subcarriers, up to 256-point QAM constellations, ≈10,000-bit LDPC codes, and channels with length $L = 256$ and average sparsity $E\{K\} = 64$. Under these conditions, we find that our scheme yields error rates that are close to genie-aided bounds, and far superior to CCS, in both low- and high-SNR regimes. Moreover, we find that the outage rate behavior of our scheme coincides with the sparse-channel capacity expression (1).

We will now place our work in context. The basic idea of using BP for joint channel-estimation and decoding (JCED) has been around for more than a decade (see, e.g., the early overview [11] and the more recent works [12,13]). The standard rules of BP specify that messages are passed among nodes of the factor graph according to the sum-product algorithm (SPA) [7]. However, since in many cases exact implementation of SPA on the JCED factor graph is impractical, SPA must be approximated, and there is considerable freedom as to how this can be done. In fact, many well known iterative estimation algorithms can be recognized as particular approximations of SPA-BP: the expectation-maximization (EM) algorithm [14], particle filtering [15], variational Bayes (or "mean-field") [16], and even steepest descent [17]. Not surprisingly, this plurality of possibilities has yielded numerous BP-based JCED designs for frequency-selective channels (e.g., [18–21]).

Our work is distinct from the existing BP-based JCED literature in that (1) we model the channel as a priori sparse (i.e., the coefficients are non-Gaussian) whereas, in all of the existing BP-based JCED work that we are aware, the channel coefficients[4] are modeled as Gaussian, and (2) we leverage a state-of-the-art BP approximation known as "relaxed BP" (RBP), which has been rigorously analyzed and shown to yield asymptotically exact posteriors (as the problem dimensions $N, L \to \infty$ and under certain technical assumptions on the mixing matrix) [8,9]. In fact, we conjecture that the success of our method is due in large part to the principled approximations used within RBP. We also note that, although we focus on the case of sparse channels, our approach would be applicable to non-sparse channels or, e.g., non-sparse channels with unknown length [18], with minor modification of the assumed channel priors.

Our paper is organized as follows. In Section 2 we detail the system model, and in Section 3 we detail our RBP-based JCED approach. In Section 4 we report the results of our simulation study, and in Section 5 we conclude.

2. System model

We assume an OFDM-based transmitter that uses a total of N subcarriers, each modulated by a QAM symbol from a 2^M-ary unit-energy constellation $\mathbb{S}$. Of these subcarriers, N_p are dedicated as pilots,[5] and the remaining $N_\text{d} \triangleq N - N_\text{p}$ are used to transmit a total of M_t training bits and $M_\text{d} \triangleq N_\text{d}M - M_\text{t}$ coded/interleaved data bits. To generate the latter, we encode M_i information bits using a rate-R coder, interleave them, and partition the resulting $M_\text{c} \triangleq M_\text{i}/R$ bits among an integer number $T \triangleq M_\text{c}/M_\text{d}$ of OFDM symbols.

[1] Note that (1) and (2) specify only that the maximum rate of reliable communication grows in linear proportion to log(SNR) according to the specified pre-log factor; the exact value of the capacity remains unspecified due to the $\mathcal{O}(1)$ term.

[2] The use of $\mathcal{O}(K\text{polylog}(L))$ pilots corresponds to the case of OFDM-based transmission, which is the case that we focus on later in this paper.

[3] In [1], a scheme that achieves the prelog factor in (1) was proposed, but it is impractical in the sense that its complexity grows exponentially with the fading-block length N.

[4] After submitting this manuscript, we became aware of the related work [22], which applies BP to JCED for flat-fading Gaussian channel coefficients and non-Gaussian interference.

[5] For our BP-based JCED, we will see in Section 4 that $(N_\text{p}, M_\text{t}) = (0, MK)$ is most effective.

Please cite this article in press as: P. Schniter, Belief-propagation-based joint channel estimation and decoding for spectrally efficient communication over unknown sparse channels, Physical Communication (2011), doi:10.1016/j.phycom.2011.09.004

The resulting scheme has a spectral efficiency of $\eta \triangleq M_{\mathrm{d}}R/N$ information bits per channel use (bpcu). It should be emphasized that our model supports both subcarriers whose QAM symbols are completely known to the receiver ("pilot subcarriers"), as well as subcarriers whose QAM symbols are only partially known to the receiver (via "training bits"). In our nomenclature, the known bits that make up a "pilot subcarrier" are distinct from the "training bits" that may be sprinkled among the "data subcarriers".

In the sequel, we use $s^{(k)} \in \mathbb{S}$ for $k \in \{1, \ldots, 2^M\}$ to denote the kth element of the QAM constellation, and $\boldsymbol{c}^{(k)} \triangleq (c_1{}^{(k)}, \ldots, c_M{}^{(k)})^{\mathsf{T}} \in \{0,1\}^M$ to denote the bits corresponding to $s^{(k)}$ as defined by the symbol mapping. Likewise, we use $s_i[t] \in \mathbb{S}$ to denote the QAM symbol transmitted on the ith subcarrier of the tth OFDM symbol and $\boldsymbol{c}_i[t] \triangleq (c_{i,1}[t], \ldots, c_{i,M}[t])^{\mathsf{T}} \in \{0,1\}^M$ to denote the (coded/interleaved or training or pilot) bits corresponding to that symbol. We then collect the NM bits that make up the tth OFDM symbol into $\boldsymbol{c}[t] \triangleq (\boldsymbol{c}_0[t], \ldots, \boldsymbol{c}_{N-1}[t])^{\mathsf{T}}$, and we collect the NMT bits that make up the entire (interleaved) codeword into $\boldsymbol{c} \triangleq (\boldsymbol{c}[1], \ldots, \boldsymbol{c}[T])^{\mathsf{T}} \in \{0,1\}^{TNM}$. The elements of $\boldsymbol{c}$ that are known a priori as pilot or training bits will be referred to as $\boldsymbol{c}_{\mathrm{pt}}$. The remainder of $\boldsymbol{c}$ is determined from the information bits $\boldsymbol{b} = (b_1, \ldots, b_{M_{\mathrm{i}}})^{\mathsf{T}}$ by coding/interleaving.

We use the standard OFDM model (see, e.g., [23]) for the received value on subcarrier i of OFDM-symbol t:

$$y_i[t] = s_i[t]z_i[t] + v_i[t], \tag{3}$$

where $z_i[t] \in \mathbb{C}$ denotes the ith subcarrier's gain and $\{v_i[t]\}$ denotes circular white Gaussian noise with variance μ^v. As usual, the subcarrier gains $\boldsymbol{z}[t] \triangleq (z_0[t], \ldots, z_{N-1}[t])^{\mathsf{T}}$ are related to the baud-spaced channel impulse response vector $\boldsymbol{x}[t] \triangleq (x_0[t], \ldots, x_{L-1}[t])^{\mathsf{T}}$ via $z_i[t] = \sum_{j=0}^{L-1} \Phi_{ij}x_j[t]$, where $\Phi_{ij} = e^{-\sqrt{-1}\frac{2\pi}{N}ij}$ can be recognized as the (i,j)th element of the N-DFT matrix Φ. Throughout, we will use j to index the lag of the impulse response. We assume that the channel is block-fading with fading interval N, so that the vectors $\{\boldsymbol{x}[t]\}_{t=1}^{T}$ are i.i.d. across t. To simplify our development of the algorithm, we assume in the sequel that $T = 1$ and drop the "$[t]$" notation for brevity. However, for the simulations in Section 4, we revert back to general T in order to facilitate the use of long LDPC codewords.

As described in Section 1, the focus of the paper is on block-Rayleigh-fading channels with sparse impulse responses $\{x_j\}$. To model sparsity, we treat the impulse response coefficients as random variables $\{X_j\}$ with the independent Bernoulli–Gaussian prior pdf.

$$p_{X_j}(x) = \lambda_j \mathcal{CN}(x; 0, \mu_j) + (1-\lambda_j)\delta(x), \tag{4}$$

where $\mathcal{CN}(x; a, b) \triangleq (\pi b)^{-1}\exp(-b^{-1}|x-a|^2)$ denotes the complex-Gaussian pdf, $\delta(\cdot)$ the Dirac delta, $\lambda_j = \Pr\{X_j \neq 0\}$ the sparsity rate, and $\mu_j = \mathrm{var}\{X_j\}$ the variance. We furthermore assume that the channel is energy-preserving with an exponential delay-power profile, so that $\mu_j = 2^{-j/L_{\mathrm{hpd}}}/(\sum_{r=0}^{L-1}\lambda_r 2^{-r/L_{\mathrm{hpd}}})$, where L_{hpd} denotes the half-power delay. For simplicity, we assume a uniform sparsity rate of $\lambda = \lambda_j \; \forall j$.

The presence of a Dirac delta in (4) indicates that we assume an "perfectly sparse" channel model. Although perfect sparsity is not expected to manifest in practice, it is frequently assumed in the literature (see, e.g., [4] and the papers cited therein). While the JCED algorithm proposed in Section 3.3 can handle generic marginal priors $p_{X_j}(x)$, we make the perfect sparsity assumption only to facilitate a direct comparison to the information theoretic result (1) from [1]. In follow-on work [24,25], we consider channel taps that are both clustered and non-perfectly sparse, as motivated by the IEEE 802.15.4a model in combination with raised-cosine pulse shapes.

3. BP-based joint channel estimation and decoding

Our goal is to infer the information bits $\boldsymbol{b}$, given the OFDM observations $\boldsymbol{y} \triangleq (y_0, \ldots, y_{N-1})^{\mathsf{T}}$ and the pilot/training bits $\boldsymbol{c}_{\mathrm{pt}}$, in the absence of channel state information. For simplicity, we assume that the channel statistics (i.e., $\{\mu^v, \lambda, L_{\mathrm{hpd}}, L\}$) are known.[6] In particular, we aim to maximize the posterior pmf $p(b_m \mid \boldsymbol{y}, \boldsymbol{c}_{\mathrm{pt}})$ of each information bit b_m. Given the model of Section 2, this posterior can be decomposed into a product of factors as follows:

$$\begin{aligned} p(b_m \mid \boldsymbol{y}, \boldsymbol{c}_{\mathrm{pt}}) &= \sum_{\boldsymbol{b}_{\setminus m}} p(\boldsymbol{b} \mid \boldsymbol{y}, \boldsymbol{c}_{\mathrm{pt}}) \propto \sum_{\boldsymbol{b}_{\setminus m}} p(\boldsymbol{y} \mid \boldsymbol{b}, \boldsymbol{c}_{\mathrm{pt}})p(\boldsymbol{b}) \qquad (5)\\ &= \int_{\boldsymbol{x}} \sum_{\boldsymbol{c}} \sum_{\boldsymbol{s}} \sum_{\boldsymbol{b}_{\setminus m}} p(\boldsymbol{y} \mid \boldsymbol{s}, \boldsymbol{x}) \\ &\quad \times p(\boldsymbol{x})p(\boldsymbol{s} \mid \boldsymbol{c})p(\boldsymbol{c} \mid \boldsymbol{b}, \boldsymbol{c}_{\mathrm{pt}})p(\boldsymbol{b}) \qquad (6)\\ &= \int_{\boldsymbol{x}} \prod_{j=0}^{L-1} p(x_j) \sum_{\boldsymbol{s}} \prod_{i=0}^{N-1} p(y_i|s_i, \boldsymbol{x}) \sum_{\boldsymbol{c}} p(s_i|\boldsymbol{c}_i) \\ &\quad \times \sum_{\boldsymbol{b}_{\setminus m}} p(\boldsymbol{c}|\boldsymbol{b}, \boldsymbol{c}_{\mathrm{pt}}) \prod_{m=1}^{M_{\mathrm{i}}} p(b_m), \qquad (7) \end{aligned}$$

where "$\propto$" denotes equality up to a scaling and where $\boldsymbol{b}_{\setminus m}$ denotes the vector $\boldsymbol{b}$ with the mth element omitted. The factorization (7) is illustrated by the *factor graph* in Fig. 1, where the round nodes represent random variables and the square nodes represent the factors of the posterior identified in (7).

3.1. Background on belief propagation

While exact evaluation of the posteriors $\{p(b_m \mid \boldsymbol{y}, \boldsymbol{c}_{\mathrm{pt}})\}_{m=1}^{M_{\mathrm{i}}}$ is computationally impractical for the problem sizes of interest, these posteriors can be approximately evaluated using *belief propagation* (BP) [7] on a loopy factor graph like the illustrated in Fig. 1. In standard BP, beliefs take the form of pdfs/pmfs that are propagated among nodes of the factor graph according to the rules of the *sum-product algorithm* (SPA):

[6] Although it remains outside the scope of this work, it should be possible to jointly estimate these statistics together with the channel and data realizations by treating them as random variables with appropriate non-informative priors and expanding the factor graph accordingly.

Please cite this article in press as: P. Schniter, Belief-propagation-based joint channel estimation and decoding for spectrally efficient communication over unknown sparse channels, Physical Communication (2011), doi:10.1016/j.phycom.2011.09.004

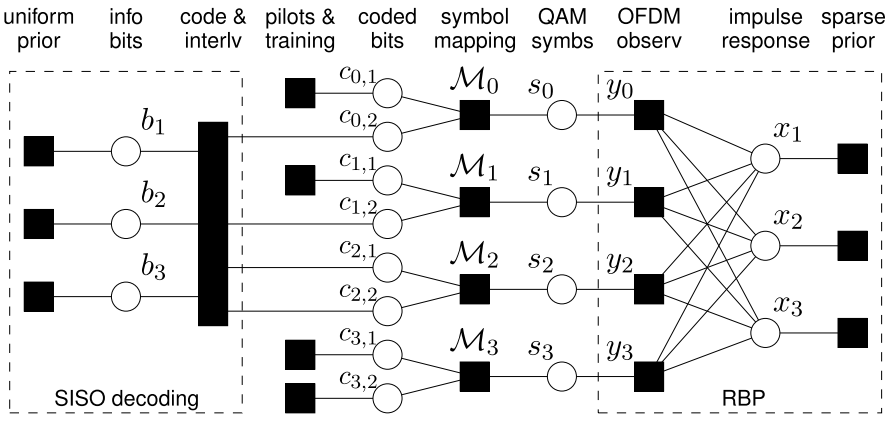

Fig. 1. The factor graph of the JCED problem for a toy example with $N = 4$ OFDM subcarriers, $M = 2$ bits per QAM symbol, $M_{\mathrm{i}} = 3$ information bits, $M_{\mathrm{t}} = 2$ training bits, $N_{\mathrm{p}} = 1$ pilot subcarriers (at subcarrier index $i = 3$), and a channel with length $L = 3$.

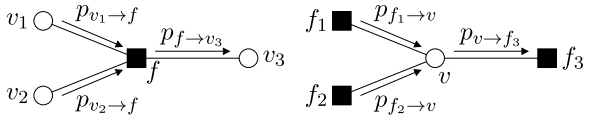

Fig. 2. Examples of belief propagation among nodes of a factor graph.

(1) If factor node $f(v_1, \ldots, v_A)$ is connected to variable nodes $\{v_a\}_{a=1}^{A}$, then the belief passed from f to v_b is

$$p_{f\to v_b}(v_b) \propto \int_{\{v_a\}_{a\neq b}} f(v_1, \ldots, v_A) \prod_{a\neq b} p_{v_a\to f}(v_a), \quad (8)$$

where $\{p_{v_a\to f}(\cdot)\}_{a\neq b}$ are the beliefs most recently passed to f from $\{v_a\}_{a\neq b}$.

(2) If variable node v is connected to factor nodes $\{f_1, \ldots, f_B\}$, then the belief passed from v to f_a is

$$p_{v\to f_a}(v) \propto \prod_{b\neq a} p_{f_b\to v}(v), \quad (9)$$

where $\{p_{f_b\to v}(\cdot)\}_{b\neq a}$ are the beliefs most recently passed to v from $\{f_b\}_{b\neq a}$.

(3) If variable node v is connected to factor nodes $\{f_1, \ldots, f_B\}$, then the posterior on v is the product of all most recently arriving beliefs, i.e.,

$$p(v) \propto \prod_{b=1}^{B} p_{f_b\to v}(v). \quad (10)$$

Fig. 2 is provided to illustrate the first two rules.

When the factor graph contains no loops, SPA-BP yields exact posteriors after only two rounds of message passing (i.e., forward and backward passes). However, with loops in the factor graph, convergence to the exact posteriors is not guaranteed, as exact inference is known to be NP-hard [26]. That said, there exist many problems to which loopy BP has been successfully applied, including inference on Markov random fields [27], multiuser detection [28,8], turbo decoding [29], LDPC decoding [10], and compressed sensing [9,30,31]. Our work not only leverages these past successes, but unites the last two through "turbo" message scheduling on a larger factor graph [32].

3.2. Background on RBP

A sub-problem of particular interest to us is the estimation of a non-Gaussian vector $\boldsymbol{x}$ that is linearly mixed to form $\boldsymbol{z} = \Phi\boldsymbol{x}$ and subsequently observed as $\boldsymbol{y}$ through componentwise non-Gaussian measurements $\{p_{Y_i|Z_i}(y_i|z_i)\}_{i=0}^{N-1}$. In our case (4) specifies the non-Gaussian prior on $\boldsymbol{x}$ and (3) yields the non-Gaussian measurement (where the non-Gaussianity results from the inherent uncertainty on data symbols s_i). This sub-problem yields the factor graph shown within the right dashed box in Fig. 1, where the nodes "y_i" represent the measurements and the rightmost nodes represent the prior on $\boldsymbol{x}$.

Building on prior multiuser detection work by Guo and Wang [8], Rangan recently proposed a so-called "relaxed BP" (RBP) scheme [9] that yields asymptotically exact posteriors as $N, L \to \infty$ (under some additional technical conditions on Φ) [9]. The main ideas behind RBP are the following. First, although the beliefs flowing leftward from the nodes $\{x_j\}$ are clearly non-Gaussian, the corresponding belief about $z_i = \sum_{j=0}^{L-1} \Phi_{ij}x_j$ can be accurately approximated as Gaussian, when L is large, using the central limit theorem. Moreover, to calculate the parameters of this distribution (i.e., its mean and variance), only the mean and variance of each x_j are needed. Thus, it suffices to pass only means and variances leftward from each x_j node. It is similarly desirable to pass only means and variances rightward from each measurement node. Although the exact rightward flowing beliefs would be non-Gaussian (due to the non-Gaussian assumption on the measurement channels $p_{Y_i|Z_i}$), RBP approximates them as Gaussian using a 2nd-order Taylor series, and passes only the resulting means and variances. A further simplification employed by RBP is to approximate the *differences* among the outgoing means/variances of each left node, and the incoming means/variances of each right node, using Taylor series. The RBP algorithm[7] is summarized in Table 1. Assuming (D1)–(D6) can be calculated efficiently (as is the case in our problem), the complexity of RBP is $\mathcal{O}(NL)$.

[7] To be precise, the RBP algorithm in Table 1 is an extension of that proposed in [9]. Table 1 handles *complex* Gaussian distributions and *non-identically distributed* signal and measurement channels.

Please cite this article in press as: P. Schniter, Belief-propagation-based joint channel estimation and decoding for spectrally efficient communication over unknown sparse channels, Physical Communication (2011), doi:10.1016/j.phycom.2011.09.004

Table 1
The RBP algorithm.

definitions:

$$p_{Z_i|Y_i}(z|y;\hat{z},\mu^z) = \frac{p_{Y_i|Z_i}(y|z)\,\mathcal{CN}(z;\hat{z},\mu^z)}{\int_{z'} p_{Y_i|Z_i}(y|z')\,\mathcal{CN}(z';\hat{z},\mu^z)} \quad \text{(D1)}$$

$$F_{\text{out},i}(y,\hat{z},\mu^z) = \int_z z\, p_{Z_i|Y_i}(z|y;\hat{z},\mu^z) \quad \text{(D2)}$$

$$\mathcal{E}_{\text{out},i}(y,\hat{z},\mu^z) = \int_z |z - F_{\text{out},i}(y,\hat{z},\mu^z)|^2\, p_{Z_i|Y_i}(z|y;\hat{z},\mu^z) \quad \text{(D3)}$$

$$p_{Q_j}(q;\hat{q},\mu^q) = \frac{p_{X_j}(q)\,\mathcal{CN}(q;\hat{q},\mu^q)}{\int_{q'} p_{X_j}(q')\,\mathcal{CN}(q';\hat{q},\mu^q)} \quad \text{(D4)}$$

$$F_{\text{in},j}(\hat{q},\mu^q) = \int_q q\, p_{Q_j}(q;\hat{q},\mu^q) \quad \text{(D5)}$$

$$\mathcal{E}_{\text{in},j}(\hat{q},\mu^q) = \int_q |q - F_{\text{in},j}(\hat{q},\mu^q)|^2\, p_{Q_j}(q;\hat{q},\mu^q) \quad \text{(D6)}$$

initialize:

$$\forall i,j: \hat{x}_{ij}[1] = \hat{x}_j[1] = \int_x x\, p_{X_j}(x) \quad \text{(I1)}$$

$$\forall j: \mu_j^x[1] = \int_x |x - \hat{x}_j[1]|^2 p_{X_j}(x) \quad \text{(I2)}$$

for $n = 1, 2, 3, \ldots$

$$\forall i: \mu_i^z[n] = \textstyle\sum_{j=0}^{L-1} |\Phi_{ij}|^2 \mu_j^x[n] \quad \text{(R1)}$$

$$\forall i: \hat{z}_i[n] = \textstyle\sum_{j=0}^{L-1} \Phi_{ij}\hat{x}_{ij}[n] \quad \text{(R2)}$$

$$\forall i,j: \hat{z}_{ij}[n] = \hat{z}_i[n] - \Phi_{ij}\hat{x}_{ij}[n] \quad \text{(R3)}$$

$$\forall i: \mu_i^e[n] = \mathcal{E}_{\text{out},i}(y_i, \hat{z}_i[n], \mu_i^z[n]) \quad \text{(R4)}$$

$$\forall i,j: \hat{e}_{ij}[n] = F_{\text{out},i}(y_i, \hat{z}_i[n], \mu_i^z[n]) - \hat{z}_{ij}[n] - \Phi_{ij}\hat{x}_{ij}[n]\mu_i^e[n]/\mu_i^z[n] \quad \text{(R5)}$$

$$\forall i: \mu_i^u[n] = \left(1 - \mu_i^e[n]/\mu_i^z[n]\right)^{-1} \mu_i^z[n] \quad \text{(R6)}$$

$$\forall i,j: \hat{u}_{ij}[n] = \left(1 - \mu_i^e[n]/\mu_i^z[n]\right)^{-1} \hat{e}_{ij}[n] \quad \text{(R7)}$$

$$\forall j: \mu_j^q[n] = \left(\textstyle\sum_{i=0}^{N-1} |\Phi_{ij}|^2/\mu_i^u[n]\right)^{-1} \quad \text{(R8)}$$

$$\forall j: \hat{q}_j[n] = \mu_j^q[n] \textstyle\sum_{i=0}^{N-1} \left(\Phi_{ij}^*\hat{u}_{ij}[n]/\mu_i^u[n]\right) \quad \text{(R9)}$$

$$\forall j: \mu_j^x[n+1] = \mathcal{E}_{\text{in},j}(\hat{q}_j[n], \mu_j^q[n]) \quad \text{(R10)}$$

$$\forall j: \hat{x}_j[n+1] = F_{\text{in},j}(\hat{q}_j[n], \mu_j^q[n]) \quad \text{(R11)}$$

$$\forall i,j: \hat{x}_{ij}[n+1] = \hat{x}_j[n+1] - \left(\Phi_{ij}^*\hat{u}_{ij}[n]/\mu_i^u[n]\right)\mu_j^x[n+1] \quad \text{(R12)}$$

end

3.3. BP-based joint channel estimation and decoding

In this section, we detail our BP-based approach to JCED, frequently referring to the factor graph in Fig. 1. Note that, since our factor graph is loopy, there exists considerable freedom in the message passing schedule. We choose to propagate beliefs from the left to the right and back again, several times, stopping as soon the beliefs have appeared to converge. Each full cycle of message passing on the overall factor graph will be referred to as a "turbo iteration". During each turbo iteration, several rounds of message passing are performed within each of the dashed boxes in Fig. 1. We refer to the iterations within the left dashed box as "SISO decoder iterations" and the iterations within the right dashed as "RBP iterations". Below, we provide details on how beliefs are calculated and propagated.

At the very start, nothing is known about the information bits, which are assumed a priori to be equally likely (i.e., $\Pr\{b_m = 1\} = \frac{1}{2}\ \forall m$). Thus, the bit beliefs that initially flow rightward out of the coding/interleaving block are uniform (i.e., $p_{c_{i,m}\to\mathcal{M}_i}(1) = \frac{1}{2}$ for all indices (i,m) corresponding to data bits). Meanwhile, the values of the pilot/training bits are known with certainty, and so $p_{c_{i,m}\to\mathcal{M}_i}(c) = 1$ for $c = c_{i,m}$.

Next, coded-bit beliefs are propagated rightward into the symbol mapping nodes. Since the symbol mapping is deterministic, the pdf factors take the form $p(s^{(k)} \mid \boldsymbol{c}^{(l)}) = \delta_{k-l}$, where $\{\delta_k\}_{k\in\mathbb{Z}}$ denotes the Kronecker delta sequence. According to the SPA, the message passed rightward from symbol mapping node "$\mathcal{M}_i$" takes the form

$$p_{\mathcal{M}_i\to s_i}(s^{(k)}) \propto \sum_{\boldsymbol{c}\in\{0,1\}^M} p(s^{(k)} \mid \boldsymbol{c}) \prod_{m=1}^{M} p_{c_{i,m}\to\mathcal{M}_i}(c_m) \quad (11)$$

$$= \prod_{m=1}^{M} p_{c_{i,m}\to\mathcal{M}_i}(c_m{}^{(k)}). \quad (12)$$

The SPA then implies that the same message is passed rightward from node s_i (i.e., $p_{\mathcal{M}_i\to s_i}(s^{(k)}) = p_{s_i\to y_i}(s^{(k)})$).

Recall, from the discussion of RBP, that the belief propagating rightward into the OFDM observation node "y_i" determines RBP's ith measurement pdf $p_{Y_i|Z_i}(y|z)$. Writing this belief as $\beta_i{}^{(k)} \triangleq p_{s_i\to y_i}(s^{(k)})$, (3) implies a Gaussian-mixture channel of the form

$$p_{Y_i|Z_i}(y|z) = \sum_{k=1}^{2^M} \beta_i{}^{(k)}\, \mathcal{CN}(y; s^{(k)}z, \mu^v). \quad (13)$$

From (13), it can be shown (see Appendix A) that the quantities (D2)–(D3) in Table 1 become

$$F_{\text{out},i}(y,\hat{z},\mu^z) = \hat{z} + \hat{e}_i(y,\hat{z},\mu^z) \quad (14)$$

$$\mathcal{E}_{\text{out},i}(y,\hat{z},\mu^z) = \sum_{k=1}^{2^M} \xi_i{}^{(k)}(y,\hat{z},\mu^z) \left(\frac{\mu^z\mu^v}{|s^{(k)}|^2\mu^z + \mu^v} + \left|\hat{e}_i(y,\hat{z},\mu^z) - \hat{e}^{(k)}(y,\hat{z},\mu^z)\right|^2 \right) \quad (15)$$

for

$$\hat{e}^{(k)}(y,\hat{z},\mu^z) \triangleq \left(\frac{y}{s^{(k)}} - \hat{z}\right) \frac{|s^{(k)}|^2\mu^z}{|s^{(k)}|^2\mu^z + \mu^v} \quad (16)$$

$$\xi_i{}^{(k)}(y,\hat{z},\mu^z) \triangleq \frac{\beta_i{}^{(k)}\mathcal{CN}(y; s^{(k)}\hat{z}, |s^{(k)}|^2\mu^z + \mu^v)}{\sum_{k'} \beta_i{}^{(k')}\mathcal{CN}(y; s^{(k')}\hat{z}, |s^{(k')}|^2\mu^z + \mu^v)} \quad (17)$$

$$\hat{e}_i(y,\hat{z},\mu^z) \triangleq \sum_{k=1}^{2^M} \xi_i{}^{(k)}(y,\hat{z},\mu^z)\, \hat{e}^{(k)}(y,\hat{z},\mu^z). \quad (18)$$

The quantities in (14)–(15) can be interpreted as follows. Given the observation $y_i = y$ and assuming the prior $z_i \sim \mathcal{CN}(\hat{z},\mu^z)$ on the subcarrier gain z_i, the quantity $F_{\text{out},i}(y,\hat{z},\mu^z)$ is the MMSE estimate of z_i, $\mathcal{E}_{\text{out},i}(y,\hat{z},\mu^z)$ is its variance, and $\{\xi_i{}^{(k)}(y,\hat{z},\mu^z)\}_{k=1}^{2^M}$ is the posterior pmf of s_i. Likewise, from (4), it can be shown (see Appendix B) that the quantities (D5)–(D6) in Table 1 take the form

$$F_{\text{in},j}(\hat{q},\mu^q) = \frac{\gamma_j(\hat{q},\mu^q)}{\alpha_j(\hat{q},\mu^q)} \quad (19)$$

$$\mathcal{E}_{\text{in},j}(\hat{q},\mu^q) = |\gamma_j(\hat{q},\mu^q)|^2 \frac{\alpha_j(\hat{q},\mu^q) - 1}{[\alpha_j(\hat{q},\mu^q)]^2} + \frac{\nu_j(\mu^q)}{\alpha_j(\hat{q},\mu^q)}, \quad (20)$$

for

$$\alpha_j(\hat{q},\mu^q) \triangleq 1 + \frac{1-\lambda_j}{\lambda_j}\frac{\mu_j}{\nu_j(\mu^q)} \exp\left(-\frac{|\gamma_j(\hat{q},\mu^q)|^2}{\nu_j(\mu^q)}\right) \quad (21)$$

Please cite this article in press as: P. Schniter, Belief-propagation-based joint channel estimation and decoding for spectrally efficient communication over unknown sparse channels, Physical Communication (2011), doi:10.1016/j.phycom.2011.09.004

$$\gamma_j(\hat{q}, \mu^q) \triangleq \frac{\nu_j(\mu^q)}{\mu^q}\hat{q} \tag{22}$$

$$\nu_j(\mu^q) \triangleq \frac{\mu^q \mu_j}{\mu^q + \mu_j}. \tag{23}$$

The quantity $F_{\text{in},j}$ from (19) can be interpreted as the MMSE estimate of the channel tap x_j given the observations $\boldsymbol{y}$ and the pilots $\boldsymbol{c}_{\text{pt}}$, and the quantity $\mathcal{E}_{\text{in},j}$ from (20) can be interpreted as its variance.

Using the quantities derived in (14)–(23), the RBP algorithm in Table 1 is iterated until convergence is detected. Doing so generates approximately conditional-mean (i.e., nonlinear MMSE) estimates $\{\hat{x}_j\}$ of the sparse-channel impulse-response coefficients $\{x_j\}$, as well as their conditional variances $\{\mu_j^x\}$, based on the observations $\{y_i\}$ and the soft symbol estimates $\{\beta_i^{(k)}\}$. Conveniently, RBP also returns (a close approximation to) the conditional-mean estimates $\{\hat{z}_i\}$ of the subchannel gains $\{z_i\}$, as well as their conditional variances $\{\mu_i^z\}$.

Before continuing, we discuss some RBP details that are specific to our JCED application. First, we notice that the condition $\mu_i^e[n] < \mu_i^z[n]$ is required to guarantee a positive value of the variance $\mu_i^u[n]$ in (R6). Intuitively, we might expect that $\mu_i^e[n] < \mu_i^z[n]$, because $\mu_i^e[n] = \mathcal{E}_{\text{out},i}(y_i, \hat{z}_i[n], \mu_i^z[n])$ is a posterior variance and $\mu_i^z[n]$ a prior variance. However, this is not necessarily the case during the first few RBP iterations, when the soft channel and symbol estimates may be inaccurate. We remedy this situation by clipping $\mu_i^e[n]$ at the value $0.99\mu^z[n]$, where 0.99 was chosen heuristically. Second, due to the DFT matrix property $|\Phi_{ij}|^2 = 1\ \forall i, j$, step (R1) in Table 1 simplifies to $\mu_i^z[n] = \mu^z[n] \triangleq \sum_{j=0}^{L-1} \mu_j^x[n]$, and (R8) simplifies to $\mu_j^q[n] = \mu^q[n] \triangleq \left(\sum_{i=0}^{N-1} 1/\mu_i^u[n]\right)^{-1}$. With these simplifications, the complexity of RBP is dominated by the computation of the elementwise matrix products $\Phi_{ij}\hat{x}_{ij}$ and $\Phi_{ij}^*\hat{u}_{ij}$, which must each be calculated once per RBP iteration, as well as three other elementwise matrix products in (R5), (R7), and (R12). Thus, RBP requires only $\approx 5NL$ multiplies per iteration.

After RBP converges, updated symbol beliefs are passed leftward out of the RBP sub-graph. According to the SPA, the belief propagating leftward from the y_i node takes the form

$$p_{s_i \leftarrow y_i}(s) \propto \int_z \mathcal{CN}(y_i; sz, \mu^v)\, \mathcal{CN}(z; \hat{z}_i, \mu_i^z) \tag{24}$$

$$= \mathcal{CN}(y_i; s\hat{z}_i, |s|^2\mu_i^z + \mu^v), \tag{25}$$

where the quantities $(\hat{z}_i, \mu_i^z)$ play the role of soft channel estimates. The SPA then implies that $p_{\mathcal{M}_i \leftarrow s_i}(s) = p_{s_i \leftarrow y_i}(s)$.

Next, beliefs are passed leftward from each symbol-mapping node $\mathcal{M}_i$ to the corresponding bit nodes $c_{i,m}$. From the SPA, these beliefs take the form

$$p_{c_{i,m} \leftarrow \mathcal{M}_i}(c) \propto \sum_{k=1}^{2^M} \sum_{\boldsymbol{c}: c_m = c} p(s^{(k)} \mid \boldsymbol{c})\, p_{\mathcal{M}_i \leftarrow s_i}(s^{(k)}) \times \prod_{m' \neq m} p_{c_{i,m'} \to \mathcal{M}_i}(c_{m'})$$

$$= \sum_{k: c_m^{(k)} = c} p_{\mathcal{M}_i \leftarrow s_i}(s^{(k)}) \frac{\prod_{m'=1}^{M} p_{c_{i,m'} \to \mathcal{M}_i}(c_{m'}^{(k)})}{p_{c_{i,m} \to \mathcal{M}_i}(c)} \tag{26}$$

$$= \frac{1}{p_{c_{i,m} \to \mathcal{M}_i}(c)} \sum_{k: c_m^{(k)} = c} p_{\mathcal{M}_i \leftarrow s_i}(s^{(k)}) p_{\mathcal{M}_i \to s_i}(s^{(k)}) \tag{27}$$

for pairs (i, m) that do not correspond to pilot/training bits. (Since the pilot/training bits are known with certainty, there is no need to update their pmfs.)

Finally, messages are passed leftward into the coding/interleaving block. Doing so is equivalent to feeding extrinsic soft bit estimates to a soft-input/soft-output (SISO) deinterleaver/decoder, which treats them as priors. Since SISO decoding is a well-studied topic (see, e.g., [10,33]) and high-performance implementations are readily available, we will not elaborate on the details here. It suffices to say that, once the extrinsic outputs of the SISO decoder have been computed, they are re-interleaved and passed rightward from the code/interleave block to begin another round of belief propagation on the overall factor graph of Fig. 1. The outer "turbo" iterations then continue until either the decoder detects no bit errors, the soft bit estimates have converged, or a maximum number of iterations has elapsed.

4. Numerical results

In this section, we present numerical results that compare our proposed BP-JCED to the CCS approach as well as to several reference schemes that act as performance upper/lower bounds.

4.1. Setup and reference schemes

The following decoupled channel-estimation and decoding (DCED) procedure was used to implement CCS. First, a LASSO[8] channel estimate $\hat{\mathbf{x}}[t]$ was generated using pilot-subcarriers. To implement LASSO, we used the celebrated SPGL1 algorithm [34] with a genie-optimized tuning parameter.[9] The frequency-domain estimate $\hat{\mathbf{z}}[t] = \Phi\hat{\mathbf{x}}[t]$ was then computed, from which the (genie-aided empirical) variance $\hat{\mu}^z[t] \triangleq \|\hat{\mathbf{z}}[t] - \boldsymbol{z}[t]\|_2^2/N$ was calculated. Using the soft channel estimate $(\hat{\mathbf{z}}[t], \hat{\mu}^z)$, leftward SPA-BP on the factor graph in Fig. 1 was performed exactly as described in Section 3.3, ensuring that the LASSO outputs were properly combined with SISO decoding. We note that, due to the two genie-aided steps, the performance attained by CCS may be somewhat optimistic. Even so, we shall see that this optimistic CCS performance remains far

[8] The criterion employed by LASSO [5] is equivalent to the one employed in "basis pursuit denoising" [6].

[9] The performance of LASSO/SPGL1 is highly dependent on the value of a tuning parameter that determines the tradeoff between the estimate's sparsity and the residual's variance. To optimize this tradeoff, for each realization, SPGL1 was invoked over a dense grid of tuning parameters, and the one that minimized NMSE (with respect to the true channel) was chosen.

Please cite this article in press as: P. Schniter, Belief-propagation-based joint channel estimation and decoding for spectrally efficient communication over unknown sparse channels, Physical Communication (2011), doi:10.1016/j.phycom.2011.09.004

below that of our BP-based JCED approach (which requires no genie-aided steps).

We now describe several reference schemes, all of which use the DCED procedure described above, but with different channel estimators. The first uses traditional linear MMSE (LMMSE) estimation. Since LMMSE does not exploit channel sparsity, it yields a performance lower-bound for any sparsity-leveraging technique. We also consider MMSE-optimal[10] pilot-aided channel estimation under the *support-aware genie* (SG), reasoning that this yields a performance upper-bound for CCS. Finally, we consider MMSE-optimal estimation under a *bit- and support-aware genie* (BSG). Here, in addition to the channel support being known, all bits (including data bits) are known and used for channel estimation. This latter reference scheme yields a performance upper-bound for *any* implementable DCED or JCED scheme, including our BP-based JCED. Remarkably, we shall see that performance of our proposed scheme is not far from that of the BSG.

For all of our results, we used irregular LDPC codes with codeword length ≈10,000 and average column weight 3, generated (and decoded) using the publicly available software [35]. Random interleaving did not seem to have an effect, and so no interleaving was employed. For bit-to-symbol mapping, we used multilevel Gray-mapping [36], noting recent work [37] that conjectures the optimality of Gray-mapping when BICM is used with a strong code. For OFDM, we used[11] $N = 1021$ subcarriers, since prime N ensures that square/tall submatrices of Φ will be full-rank. As described in the sequel, we tested various combinations of pilot subcarriers N_p and interspersed training bits M_t. The N_p pilot subcarriers were spaced uniformly and modulated with QAM symbols chosen uniformly at random. The M_t training bits were placed at the most significant bits (MSBs) of uniformly spaced data subcarriers with values chosen uniformly at random.

Unless otherwise specified, we used length $L = 256$ channels with sparsity rate $\lambda = 1/4$, yielding $E\{K\} = \lambda N = 64$ non-zero taps on average. All results are averaged over $T = 100$ OFDM symbols.

4.2. NMSE *and* BER *versus the number of pilot subcarriers*

Fig. 3 plots channel estimation normalized mean-squared error $\mathsf{NMSE} \triangleq \|\hat{\mathbf{x}}[t] - \boldsymbol{x}[t]\|_2^2 / \|\boldsymbol{x}[t]\|_2^2$ versus the pilot-to-sparsity ratio N_p/K at $\mathsf{SNR} = 20$ dB. As expected, CCS's NMSE falls between that of LMMSE and SG estimators, and all three decrease monotonically with N_p/K. Even after a single turbo iteration, BP-JCED significantly outperforms CCS, and – perhaps surprisingly – the SG (when $N_\mathrm{p}/K \geq 3$). The reason for this latter behavior is that, while the SG uses only the N_p pilot subcarriers, BP-JCED uses all N subcarriers, which yields improved performance even though the $N_\mathrm{d} = N - N_\mathrm{p}$ data symbols are known

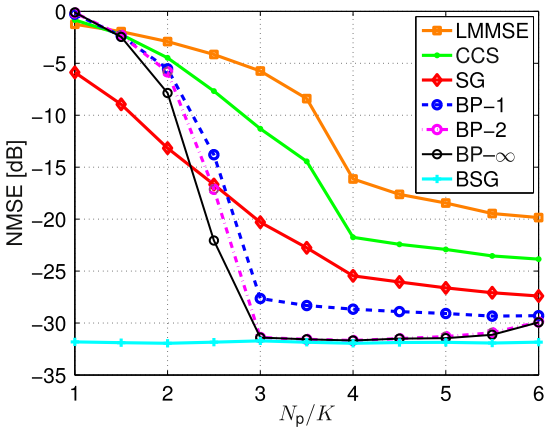

Fig. 3. Channel estimation NMSE versus pilot-to-sparsity ratio N_p/K, for $\mathsf{SNR} = 20$ dB, $M_\mathrm{t} = 0$ training bits, $\eta = 3$ bpcu, and 64-QAM.

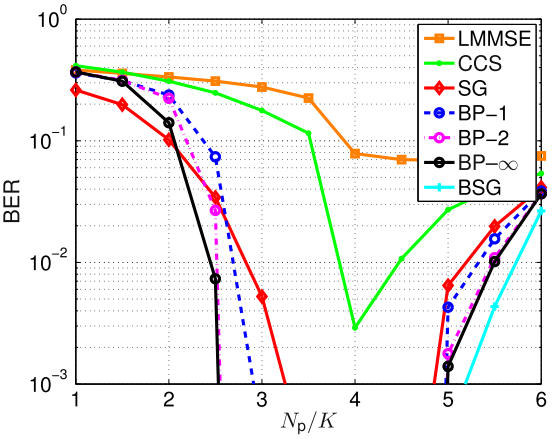

Fig. 4. BER versus pilot ratio N_p/K, for $\mathsf{SNR} = 20$ dB, $M_\mathrm{t} = 0$ training bits, $\eta = 3$ bpcu, and 64-QAM.

with very little certainty during the first turbo iteration. Fig. 3 indicates that, after only 2 turbo iterations, BP-JCED learns the data symbols well enough to estimate the channel nearly as well as the BSG (which knows the data symbols perfectly). The fact that BP-JCED can generate channel estimates that are nearly as good as BSG's support-aware estimates attests to the near-optimal compressive estimation abilities of RBP.

Fig. 4 plots bit error rate (BER) versus the pilot ratio N_p/K at $\mathsf{SNR} = 20$ dB and a fixed spectral efficiency of $\eta = 3$ bpcu. The curves exhibit a "notched" shape because, as N_p increases, the code rate R must decrease to maintain a fixed value of spectral efficiency η. Thus, while an increase in N_p can make channel estimation easier, the reduction in R makes data decoding more difficult. For CCS, Fig. 4 indicates that $N_\mathrm{p} = 4K = L$ is optimal. The SG and BP-JCED curves show a similar notch-like shape, although their notches are much wider. Finally, the degradation of BP-JCED's data-bit estimates at large N_p/K explains the degradation of its channel estimates, as seen in Fig. 3, since, with JCED, channel estimation is data-directed.

It is interesting to notice that Fig. 4 shows the optimal CCS pilot insertion rate to be the "Nyquist" rate of $N_\mathrm{p} = L$,

[10] When the sparse-channel support is known, the non-zero channel coefficients follow a Gaussian prior, and MMSE-optimal estimates can be calculated linearly.

[11] Experiments with non-prime $N = 1024$ showed a slight degradation of performance.

Please cite this article in press as: P. Schniter, Belief-propagation-based joint channel estimation and decoding for spectrally efficient communication over unknown sparse channels, Physical Communication (2011), doi:10.1016/j.phycom.2011.09.004

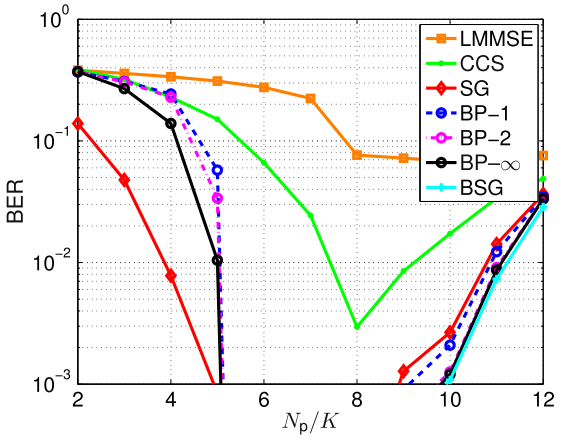

Fig. 5. BER versus pilot ratio N_p/K, for SNR $= 20$ dB, $M_t = 0$ training bits, $\eta = 3$ bpcu, and 64-QAM. The channel used here had the sparsity rate $\lambda = E\{K\}/L = 1/8$, which is half the value used in all other experiments.

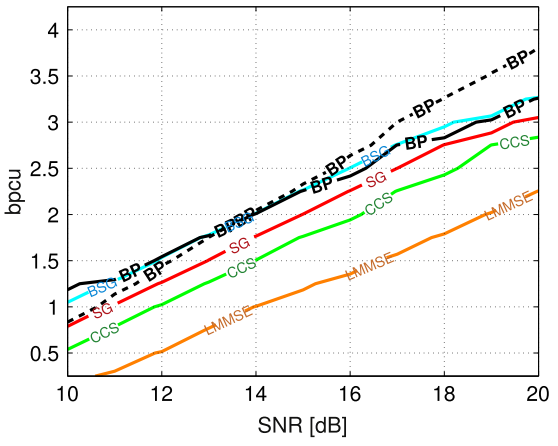

Fig. 6. BER $= 0.001$-achieving spectral efficiency $\eta_{0.001}$ versus SNR. The solid traces used $N_p/K = 4$, $M_t = 0$, and 64-QAM, while the dashed trace used $N_p = 0$, $M_t = MK$, and 256-QAM.

since, at this pilot rate, CCS is not actually "compressed". To further investigate this behavior, we repeated the experiment using a channel with half the number of active coefficients (i.e., $\lambda = E\{K\}/L = 1/8$) and report the results in Fig. 5. Remarkably, we find the same behavior: CCS again performs best when pilots are inserted at the Nyquist rate of $N_p = L$. In fact, we repeated this experiment with dozens of other arbitrary combinations of $(N, L, \lambda, \text{SNR}, \eta, M)$, and always found exactly the same behavior. Our empirical evidence suggests that, generally speaking, *decoupled sparse-channel estimation and data decoding works best when pilots are inserted at the Nyquist rate*, at least for OFDM signaling under uniform subcarrier power allocation.[12]

4.3. Outage rate and the importance of bit-level training

Fig. 6 plots $\eta_{0.001}$ versus SNR, where $\eta_{0.001}$ denotes the spectral efficiency (in bpcu) yielding BER $= 0.001$. The solid-line traces correspond to $N_p = 4K = L$ pilots, $M_t = 0$ training bits, and 64-QAM, as suggested by Fig. 4. These solid-line traces all display the anticipated high-SNR scaling law $(1 - N_p/N)\log_2(\text{SNR}) + \mathcal{O}(1)$, differing only in the $\mathcal{O}(1)$ offset term. While, for this setup, we are glad to see BP-JCED performing on par with BSG, neither attains the desired channel-capacity prelog-factor of $(1 - K/N) = 15/16$. It turns out that this shortcoming is due to the choice $(N_p, M_t) = (L, 0)$, which was chosen on behalf of CCS (and not BP-JCED).

To find the optimal choice of (N_p, M_t) for BP-JCED, we constructed the BER plot Fig. 7. There we see that BP-JCED performs best with $(N_p, M_t) = (0, MK)$, at least in the high-SNR regime. Note that the total number of pilot/training bits used when $(N_p, M_t) = (0, MK)$ is equivalent to K degrees-of-freedom per fading block, consistent with the channel-capacity prelog factor. We then evaluated the outage rate of this scheme (with 256-QAM), obtaining the dashed $\eta_{0.001}$-vs-SNR trace in Fig. 6, which – remarkably – exhibits the desired prelog-factor of $(1 - K/N)$.

[12] It would be interesting to see if this behavior persists when the pilot- versus data-subcarrier power allocation is optimized. Such an optimization, however, remains outside the scope of this manuscript.

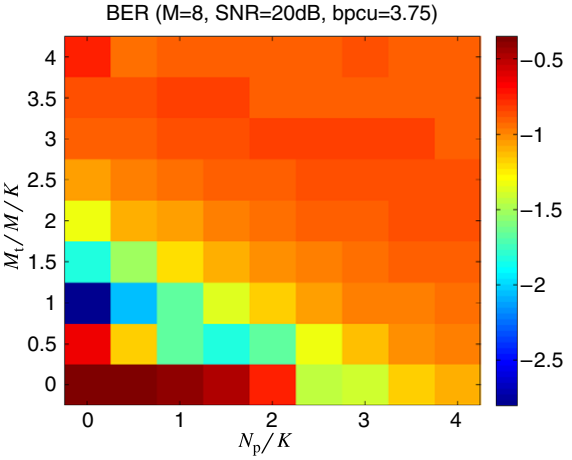

Fig. 7. $\log_{10}(\text{BER})$ versus various combinations of pilot and training rate, for SNR $= 20$ dB, $\eta = 3.75$ bpcu, and 256-QAM.

Fig. 8 plots BER versus $E_b/N_o \triangleq \text{SNR}/\eta$ over a much lower range of SNR. As stated earlier, experiments confirmed that CCS favors $(N_p, M_t) = (L, 0)$ in the low-SNR regime, and so this configuration was used to keep CCS competitive, while being potentially suboptimal for BP-JCED. Still, we see from Fig. 8 that BP-JCED, after only two turbo iterations, beats CCS by 1.8 dB and remains only 0.8 dB away from the BSG.

5. Conclusion

In this work, we presented a novel approach to joint channel estimation and decoding (JCED) for spectrally efficient communication over channels with possibly sparse impulse responses. For this, we assumed a pilot-aided transmission scheme that combines bit interleaved coded modulation (BICM) with orthogonal frequency division multiplexing (OFDM). Our JCED scheme is based on belief propagation (BP) over a loopy factor graph, where our BP implementation uses very efficient approximations of the sum-product algorithm recently proposed under the guise

Please cite this article in press as: P. Schniter, Belief-propagation-based joint channel estimation and decoding for spectrally efficient communication over unknown sparse channels, Physical Communication (2011), doi:10.1016/j.phycom.2011.09.004

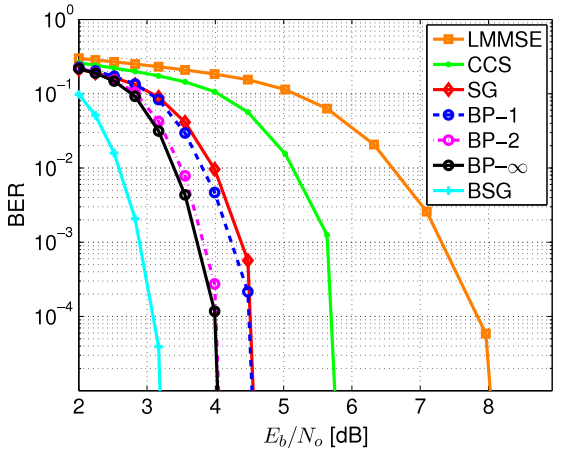

Fig. 8. BER versus $E_b/N_o(\triangleq \text{SNR}/\eta)$, for $N_p/K = 4$, $M_t = 0$, $\eta = 0.5$ bpcu, and 4-QAM.

of relaxed belief propagation (RBP) [8,38] and soft-input soft-output decoding. Because our JCED scheme requires only $\approx 5NL$ multiplications per RBP iteration, we can handle long impulse responses, large numbers of OFDM subcarriers, and large constellations. Numerical experiments conducted using $N = 1021$ subcarriers, up to 256-point QAM constellations, ≈10,000-bit LDPC codes, and channels with length $L = 256$ and average sparsity $E\{K\} = 64$, showed that the BER of BP-JCED is close to genie-aided bounds and much better than the BER of the LASSO-based "compressed channel sensing" (CCS) approach, where sparse channel estimation is decoupled from data decoding. Moreover, the outage rates observed for BP-JCED exhibit the sparse-channel capacity pre-log factor $(1-K/N)$, which is impossible to reach using CCS.

Acknowledgments

This work was supported in part by the National Science Foundation grant CCF-1018368 and DARPA/ONR grant N66001-10-1-4090.

Portions of this work were presented at the Asilomar 2010 Conference on Signals, Systems, and Computers.

Appendix A. Derivation of RBP Quantities $F_{\text{out},i}$ and $E_{\text{out},i}$

In this Appendix A, we derive the RBP quantities $F_{\text{out},i}(y,\hat{z},\mu^z)$ and $\mathcal{E}_{\text{out},i}(y,\hat{z},\mu^z)$ given in (14)–(18). From (D1)–(D2), we have that

$$F_{\text{out},i}(y,\hat{z},\mu^z) = \frac{1}{p_{Y_i}(y)}\int_z z\, p_{Y_i|Z_i}(y|z)\,\mathcal{CN}(z;\hat{z},\mu^z), \quad (28)$$

where $p_{Y_i}(y) \triangleq \int_z p_{Y_i|Z_i}(y|z)\mathcal{CN}(z;\hat{z},\mu^z)$. From (13), we rewrite $p_{Y_i|Z_i}(y|z)$ as

$$p_{Y_i|Z_i}(y|z) = \sum_{k=1}^{2^M}\frac{\beta_i^{(k)}}{s^{(k)}}\,\mathcal{CN}\left(z;\frac{y}{s^{(k)}},\frac{\mu^v}{|s^{(k)}|^2}\right), \quad (29)$$

so that

$$\int_z z\, p_{Y_i|Z_i}(y|z)\mathcal{CN}(z;\hat{z},\mu^z) = \sum_{k=1}^{2^M}\frac{\beta_i^{(k)}}{s^{(k)}}\int_z z\,\mathcal{CN}\left(z;\frac{y}{s^{(k)}},\frac{\mu^v}{|s^{(k)}|^2}\right)\mathcal{CN}(z;\hat{z},\mu^z) \quad (30)$$

$$p_{Y_i}(y) = \sum_{k=1}^{2^M}\frac{\beta_i^{(k)}}{s^{(k)}}\int_z \mathcal{CN}\left(z;\frac{y}{s^{(k)}},\frac{\mu^v}{|s^{(k)}|^2}\right)\mathcal{CN}(z;\hat{z},\mu^z). \quad (31)$$

Using the property that

$$\mathcal{CN}(x;\hat{\theta},\mu^\theta)\mathcal{CN}(x;\hat{\phi},\mu^\phi) = \mathcal{CN}\left(x;\frac{\hat{\theta}/\mu^\theta+\hat{\phi}/\mu^\phi}{1/\mu^\theta+1/\mu^\phi},\frac{1}{1/\mu^\theta+1/\mu^\phi}\right) \times \mathcal{CN}(0;\hat{\theta}-\hat{\phi},\mu^\theta+\mu^\phi), \quad (32)$$

we can rewrite

$$\int_z z\, p_{Y_i|Z_i}(y|z)\,\mathcal{CN}(z;\hat{z},\mu^z) = \sum_{k=1}^{2^M}\frac{\beta_i^{(k)}}{s^{(k)}}\mathcal{CN}\left(0;\frac{y_i}{s}-\hat{z},\frac{\mu^v}{|s^{(k)}|^2}+\mu^z\right) \times \int_z z\,\mathcal{CN}\left(z;\frac{\frac{y}{s^{(k)}}\frac{|s^{(k)}|^2}{\mu^v}+\frac{\hat{z}}{\mu^z}}{\frac{|s^{(k)}|^2}{\mu^v}+\frac{1}{\mu^z}},\frac{1}{\frac{|s^{(k)}|^2}{\mu^v}+\frac{1}{\mu^z}}\right) \quad (33)$$

$$= \sum_{k=1}^{2^M}\frac{\beta_i^{(k)}}{s^{(k)}} \times \mathcal{CN}\left(\frac{y_i}{s};\hat{z},\frac{\mu^v}{|s^{(k)}|^2}+\mu^z\right)\frac{\frac{y}{s^{(k)}}\frac{|s^{(k)}|^2}{\mu^v}+\frac{\hat{z}}{\mu^z}}{\frac{|s^{(k)}|^2}{\mu^v}+\frac{1}{\mu^z}} \quad (34)$$

$$= \sum_{k=1}^{2^M}\beta_i{}^{(k)}\mathcal{CN}\left(y_i;s^{(k)}\hat{z},|s^{(k)}|^2\mu^z+\mu^v\right) \times \left(\underbrace{\left(\frac{y}{s^{(k)}}-\hat{z}\right)\frac{|s^{(k)}|^2\mu^z}{|s^{(k)}|^2\mu^z+\mu^v}}_{\triangleq\hat{e}^{(k)}(y,\hat{z},\mu^z)}+\hat{z}\right) \quad (35)$$

and, using the same procedure, we get

$$p_{Y_i}(y) = \sum_{k=1}^{2^M}\beta_i{}^{(k)}\mathcal{CN}\left(y_i;s^{(k)}\hat{z},|s^{(k)}|^2\mu^z+\mu^v\right). \quad (36)$$

Finally, with $\xi_i^{(k)}(y,\hat{z},\mu^z)$ defined in (17), Eqs. (28), (35) and (36) combine to give

$$F_{\text{out},i}(y,\hat{z},\mu^z) = \sum_{k=1}^{2^M}\xi_i{}^{(k)}(y,\hat{z},\mu^z)\left(\hat{e}^{(k)}(y,\hat{z},\mu^z)+\hat{z}\right), \quad (37)$$

from which (14) follows immediately.

Please cite this article in press as: P. Schniter, Belief-propagation-based joint channel estimation and decoding for spectrally efficient communication over unknown sparse channels, Physical Communication (2011), doi:10.1016/j.phycom.2011.09.004

From (D1) to (D3), we have that

$$\mathcal{E}_{\text{out},i}(y,\hat{z},\mu^z) = \frac{\int_z |z-F_{\text{out},i}|^2\, p_{Y_i|Z_i}(y|z)\, \mathcal{CN}(z;\hat{z},\mu^z)}{p_{Y_i}(y)}. \tag{38}$$

Similar to (33), we can write

$$\begin{aligned}
&\int_z |z-F_{\text{out},i}|^2\, p_{Y_i|Z_i}(y|z)\, \mathcal{CN}(z;\hat{z},\mu^z) \\
&= \sum_{k=1}^{2^M} \frac{\beta_i^{(k)}}{s^{(k)}} \mathcal{CN}\left(0; \frac{y_i}{s} - \hat{z}, \frac{\mu^v}{|s^{(k)}|^2} + \mu^z\right) \\
&\quad \times \int_z |z-F_{\text{out},i}|^2 \\
&\quad \times \mathcal{CN}\left(z; \frac{\frac{y}{s^{(k)}}\frac{|s^{(k)}|^2}{\mu^v} + \frac{\hat{z}}{\mu^z}}{\frac{|s^{(k)}|^2}{\mu^v} + \frac{1}{\mu^z}}, \frac{1}{\frac{|s^{(k)}|^2}{\mu^v} + \frac{1}{\mu^z}}\right).
\end{aligned} \tag{39}$$

Then, using the change-of-variable $\tilde{z} \triangleq z - F_{\text{out},i}$, and absorbing the $s^{(k)}$ terms as we did in (35), we get

$$\begin{aligned}
&\int_z |z-F_{\text{out},i}|^2\, p_{Y_i|Z_i}(y|z)\, \mathcal{CN}(z;\hat{z},\mu^z) \\
&= \sum_{k=1}^{2^M} \beta_i^{(k)} \mathcal{CN}\left(y_i; s^{(k)}\hat{z}, |s^{(k)}|^2\mu^z + \mu^v\right) \int_{\tilde{z}} |\tilde{z}|^2 \\
&\quad \times \mathcal{CN}\left(\tilde{z}; \hat{e}^{(k)} + \underbrace{\hat{z} - F_{\text{out},i}}_{=-\hat{e}_i}, \frac{\mu^v\mu^z}{|s^{(k)}|^2\mu^z + \mu^v}\right)
\end{aligned} \tag{40}$$

$$\begin{aligned}
&= \sum_{k=1}^{2^M} \beta_i^{(k)} \mathcal{CN}\left(y_i; s^{(k)}\hat{z}, |s^{(k)}|^2\mu^z + \mu^v\right) \\
&\quad \times \left(|\hat{e}^{(k)} - \hat{e}_i|^2 + \frac{\mu^v\mu^z}{|s^{(k)}|^2\mu^z + \mu^v}\right).
\end{aligned} \tag{41}$$

Finally, using $\xi_i^{(k)}(y,\hat{z},\mu^z)$ defined in (17), Eqs. (36), (38) and (41) combine to give the expression for $\mathcal{E}_{\text{out},i}(y,\hat{z},\mu^z)$ given in (15).

Appendix B. Derivation of RBP quantities $F_{\text{in},j}$ and $\mathcal{E}_{\text{in},j}$

In this Appendix B, we derive the RBP quantities $F_{\text{in},j}(\hat{q},\mu^q)$ and $\mathcal{E}_{\text{in},j}(\hat{q},\mu^q)$ given in (19)–(23).

From (D4)–(D6), we note that $F_{\text{in},j}(\hat{q},\mu^q)$ and $\mathcal{E}_{\text{in},j}(\hat{q},\mu^q)$ are the mean and variance, respectively, of the pdf

$$\frac{1}{Z_j} p_{X_j}(q)\, \mathcal{CN}(q;\hat{q},\mu^q), \tag{42}$$

where $Z_j = \int_q p_{X_j}(q)\, \mathcal{CN}(q;\hat{q},\mu^q)$. Using (32) together with the definition of $p_{X_j}(.)$ from (4), we find that

$$\begin{aligned}
p_{X_j}(q)\, \mathcal{CN}(q;\hat{q},\mu^q) &= \lambda_j \mathcal{CN}(q;0,\mu_j)\, \mathcal{CN}(q;\hat{q},\mu^q) \\
&\quad + (1-\lambda_j)\delta(q)\, \mathcal{CN}(q;\hat{q},\mu^q)
\end{aligned} \tag{43}$$

$$\begin{aligned}
&= \lambda_j \mathcal{CN}(0;-\hat{q},\mu_j+\mu^q) \\
&\quad \times \mathcal{CN}\left(q; \frac{\hat{q}/\mu^q}{1/\mu_j + 1/\mu^q}, \frac{1}{1/\mu_j + 1/\mu^q}\right) \\
&\quad + (1-\lambda_j)\mathcal{CN}(0;\hat{q},\mu^q)\delta(q)
\end{aligned} \tag{44}$$

$$\begin{aligned}
&= \lambda_j \mathcal{CN}(\hat{q};0,\mu_j+\mu^q) \\
&\quad \times \mathcal{CN}\left(q; \frac{\hat{q}}{\mu^q}\frac{\mu^q\mu_j}{\mu^q+\mu_j}, \frac{\mu^q\mu_j}{\mu^q+\mu_j}\right) \\
&\quad + (1-\lambda_j)\mathcal{CN}(\hat{q};0,\mu^q)\delta(q),
\end{aligned} \tag{45}$$

which implies that

$$Z_j = \lambda_j \mathcal{CN}(\hat{q};0,\mu_j+\mu^q) + (1-\lambda_j)\mathcal{CN}(\hat{q};0,\mu^q). \tag{46}$$

Thus, the mean obeys

$$F_{\text{in},j}(\hat{q},\mu^q) = \frac{1}{Z_j}\int_q q\, p_{X_j}(q)\, \mathcal{CN}(q;\hat{q},\mu^q) \tag{47}$$

$$= \underbrace{\frac{\lambda_j}{Z_j}\mathcal{CN}(\hat{q};0,\mu_j+\mu^q)}_{=1/\alpha_j(\hat{q},\mu^q)} \underbrace{\frac{\hat{q}}{\mu^q}\frac{\mu^q\mu_j}{\mu^q+\mu_j}}_{=\gamma_j(\hat{q},\mu^q)}. \tag{48}$$

Expression (19) then follows directly from (48).

Since, for the pdf in (42), $F_{\text{in},j}$ is the mean and $\mathcal{E}_{\text{in},j}$ is the variance, we can write

$$\mathcal{E}_{\text{in},j}(\hat{q},\mu^q) = \frac{1}{Z_j}\int_q |q-F_{\text{in},j}|^2\, p_{X_j}(q)\, \mathcal{CN}(q;\hat{q},\mu^q) \tag{49}$$

$$= \frac{1}{Z_j}\int_q |q|^2\, p_{X_j}(q)\, \mathcal{CN}(q;\hat{q},\mu^q) - |F_{\text{in},j}|^2 \tag{50}$$

$$\begin{aligned}
&= \frac{1}{Z_j}\int_q |q|^2 \Big(\lambda_j \mathcal{CN}(\hat{q};0,\mu_j+\mu^q) \\
&\quad \times \mathcal{CN}\left(q; \frac{\hat{q}}{\mu^q}\frac{\mu^q\mu_j}{\mu^q+\mu_j}, \frac{\mu^q\mu_j}{\mu^q+\mu_j}\right) \\
&\quad + (1-\lambda_j)\mathcal{CN}(\hat{q};0,\mu^q)\delta(q)\Big) - |F_{\text{in},j}|^2
\end{aligned} \tag{51}$$

$$\begin{aligned}
&= \underbrace{\frac{\lambda_j}{Z_j}\mathcal{CN}(\hat{q};0,\mu_j+\mu^q)}_{=1/\alpha_j} \\
&\quad \times \left(\left|\underbrace{\frac{\hat{q}}{\mu^q}\frac{\mu^q\mu_j}{\mu^q+\mu_j}}_{=\gamma_j}\right|^2 + \underbrace{\frac{\mu^q\mu_j}{\mu^q+\mu_j}}_{=\nu_j}\right) - |F_{\text{in},j}|^2
\end{aligned} \tag{52}$$

$$= \frac{1}{\alpha_j}\left(|\gamma_j|^2 + \nu_j\right) - \frac{1}{\alpha_j^2}|\gamma_j|^2. \tag{53}$$

Expression (20) then follows by rearranging (53).

References

[1] A.P. Kannu, P. Schniter, On communication over unknown sparse frequency-selective block-fading channels, IEEE Trans. Inform. Theory, http://arXiv.org/abs/1006.1548, June 2010 (in press).
[2] H. Vikalo, B. Hassibi, B. Hochwald, T. Kailath, On the capacity of frequency-selective channels in training-based transmission schemes, IEEE Trans. Signal Process. (2004) 2572–2583.
[3] L. Tong, B.M. Sadler, M. Dong, Pilot-assisted wireless transmissions, IEEE Signal Process. Mag. 21 (2004) 12–25.
[4] W.U. Bajwa, J. Haupt, A.M. Sayeed, R. Nowak, Compressed channel sensing: a new approach to estimating sparse multipath channels, Proc. IEEE 98 (2010) 1058–1076.

Please cite this article in press as: P. Schniter, Belief-propagation-based joint channel estimation and decoding for spectrally efficient communication over unknown sparse channels, Physical Communication (2011), doi:10.1016/j.phycom.2011.09.004

[5] R. Tibshirani, Regression shrinkage and selection via the lasso, J. Roy. Statist. Soc. B 58 (1) (1996) 267–288.
[6] S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM J. Sci. Comput. 20 (1) (1998) 33–61.
[7] J. Pearl, Probabilistic Reasoning in Intelligent Systems, Morgan Kaufman, San Mateo, CA, 1988.
[8] D. Guo, C.-C. Wang, Random sparse linear systems observed via arbitrary channels: a decoupling principle, in: Proc. IEEE Int. Symposium Inform. Theory, Nice, France, June 2007, pp. 946–950.
[9] S. Rangan, Estimation with random linear mixing, belief propagation and compressed sensing, http://arXiv.org/abs/1001.2228v2, May 2010.
[10] D.J.C. MacKay, Information Theory, Inference, and Learning Algorithms, Cambridge University Press, New York, 2003.
[11] A.P. Worthen, W.E. Stark, Unified design of iterative receivers using factor graphs, IEEE Trans. Inform. Theory 47 (2001) 843–849.
[12] X. Jin, A.W. Eckford, T.E. Fuja, LDPC codes for non-coherent block fading channels with correlation: analysis and design, IEEE Trans. Commun. 56 (1) (2008) 70–80.
[13] Q. Guo, L. Ping, LMMSE turbo equalization based on factor graphs, IEEE J. Sel. Areas Commun. 26 (2008) 311–319.
[14] J. Dauwels, S. Korl, H.-A. Loeliger, Expectation maximization as message passing, in: Proc. IEEE Int. Symposium Inform. Theory, Adelaide, SA, Sep. 2005, pp. 425–429.
[15] J. Dauwels, S. Korl, H.-A. Loeliger, Particle methods as message passing, in: Proc. IEEE Int. Symposium Inform. Theory, Seattle, WA, Jul. 2006, pp. 2052–2056.
[16] J. Dauwels, On variational message passing on factor graphs, in: Proc. IEEE Int. Symposium Inform. Theory, Nice, France, June 2007, pp. 2546–2550.
[17] J. Dauwels, S. Korl, H.-A. Loeliger, Steepest descent as message passing, in: Proc. Inform. Theory Workshop, Awaji, Japan, Oct. 2005, pp. 42–46.
[18] C. Novak, G. Matz, F. Hlawatsch, Factor graph based design of an OFDM-IDMA receiver performing joint data detection, channel estimation, and channel length selection, in: Proc. IEEE Int. Conf. Acoust. Speech & Signal Process., Taipei, Taiwan, Apr. 2009, pp. 2433–2436.
[19] Y. Liu, L. Brunel, J. Boutros, Joint channel estimation and decoding using Gaussian approximation in a factor graph over multipath channel, in: Proc. IEEE Int. Symposium Pers. Indoor Mobile Radio Commun., Tokyo, Japan, Sept. 2009, pp. 3164–3168.
[20] C. Kneivel, Z. Shi, P.A. Hoeher, G. Auer, 2D graph-based soft channel estimation for MIMO-OFDM, in: Proc. IEEE Int. Conf. Commun., Cape Town, Jul. 2010, pp. 1–5.
[21] G.E. Kirkelund, C.N. Manchón, L.P.B. Christensen, E. Riegler, B.H. Fleury, Variational message-passing for joint channel estimation and decoding in MIMO-OFDM, in: Proc. IEEE Global Telecommun. Conf., Miami, FL, Dec. 2010.
[22] Y. Zhu, D. Guo, M.L. Honig, A message-passing approach for joint channel estimation, interference mitigation, and decoding, IEEE Trans. Wireless Commun. 8 (2009) 6008–6018.
[23] L.J. Cimini Jr., Analysis and simulation of a digital mobile radio channel using orthogonal frequency division multiplexing, IEEE Trans. Commun. 33 (1985) 665–765.
[24] P. Schniter, A message-passing receiver for BICM-OFDM over unknown clustered-sparse channels, in: Proc. IEEE Workshop Signal Process. Adv. Wireless Commun., San Francisco, CA, June 2011.
[25] P. Schniter, A message-passing receiver for BICM-OFDM over unknown clustered-sparse channels, IEEE J. Sel. Top. Signal Process. http://arXiv.org/abs/1101.4724, Jan. 2011 (in press).
[26] G.F. Cooper, The computational complexity of probabilistic inference using Bayesian belief networks, Artificial Intelligence 42 (1990).
[27] W.T. Freeman, E.C. Pasztor, O.T. Carmichael, Learning low-level vision, Intl. J. Comput. Vision 40 (2000) 25–47.
[28] J. Boutros, G. Caire, Iterative multiuser joint decoding: unified framework and asymptotic analysis, IEEE Trans. Inform. Theory 48 (2002) 1772–1793.
[29] R.J. McEliece, D.J.C. MacKay, J.-F. Cheng, Turbo decoding as an instance of Pearl's belief propagation algorithm, IEEE J. Sel. Areas Commun. 16 (1998) 140–152.
[30] D. Baron, S. Sarvotham, R.G. Baraniuk, Bayesian compressive sensing via belief propagation, IEEE Trans. Signal Process. 58 (2010) 2269–2280.
[31] D.L. Donoho, A. Maleki, A. Montanari, Message passing algorithms for compressed sensing, Proc. Natl. Acad. Sci. 106 (2009) 18914–18919.
[32] P. Schniter, Turbo reconstruction of structured sparse signals, in: Proc. Conf. Inform. Science & Syst., Princeton, NJ, Mar. 2010.
[33] T.J. Richardson, R.L. Urbanke, Modern Coding Theory, Cambridge University Press, New York, 2009.
[34] E. van den Berg, M.P. Friedlander, Probing the Pareto frontier for basis pursuit solutions, SIAM J. Sci. Comput. 31 (2) (2008) 890–912.
[35] I. Kozintsev, Matlab programs for encoding and decoding of LDPC codes over GF(2^m). http://www.kozintsev.net/soft.html.
[36] Y.L.C. de Jong, T.J. Willink, Iterative tree search detection for MIMO wireless systems, IEEE Trans. Commun. 53 (2005) 930–935.
[37] M. Samuel, M. Barsoum, M.P. Fitz, On the suitability of Gray bit mappings to outer channel codes in iteratively decoded BICM, in: Proc. Asilomar Conf. Signals Syst. Comput., Pacific Grove, CA, Nov. 2009, pp. 982–985.
[38] S. Rangan, Generalized approximate message passing for estimation with random linear mixing, http://arXiv.org/abs/1010.5141, Oct. 2010.

Philip Schniter received the B.S. and M.S. degrees in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign in 1992 and 1993, respectively.

From 1993 to 1996 he was employed by Tektronix Inc. in Beaverton, OR as a systems engineer.

In 2000, he received the Ph.D. degree in Electrical Engineering from Cornell University in Ithaca, NY.

Subsequently, he joined the Department of Electrical and Computer Engineering at The Ohio State University in Columbus, OH, where he is now an Associate Professor and a member of the Information Processing Systems (IPS) Lab.

In 2008–2009, he held visiting positions at EURECOM (Sophia-Antipolis, France) and Supelec (Gif-sur-Yvette, France).

Dr. Schniter's current research interests focus on signal processing, estimation, and communication.

Please cite this article in press as: P. Schniter, Belief-propagation-based joint channel estimation and decoding for spectrally efficient communication over unknown sparse channels, Physical Communication (2011), doi:10.1016/j.phycom.2011.09.004