A MESSAGE-PASSING RECEIVER FOR BICM-OFDM OVER UNKNOWN CLUSTERED-SPARSE CHANNELS

Philip Schniter

Dept. ECE, The Ohio State University, Columbus OH 43210. (schniter@ece.osu.edu)

ABSTRACT

We propose a factor-graph-based approach to joint channel-estimation-and-decoding of bit-interleaved coded orthogonal frequency division multiplexing (BICM-OFDM). In contrast to existing designs, ours is capable of exploiting not only sparsity in sampled channel taps but also clustering among the large taps, behaviors which are known to manifest at larger communication bandwidths. In order to exploit these channel-tap structures, we adopt a two-state Gaussian mixture prior in conjunction with a Markov model on the hidden state. For loopy belief propagation, we exploit a "generalized approximate message passing" algorithm recently developed in the context of compressed sensing, and show that it can be successfully coupled with soft-input soft-output decoding, as well as hidden Markov inference. For N subcarriers and M bits per subcarrier (and any channel length $L < N$), our scheme has a computational complexity of only $\mathcal{O}(N \log_2 N + N2^M)$. Numerical experiments using IEEE 802.15.4a channels show that our scheme yields BER performance within 1 dB of the known-channel bound and 4 dB better than decoupled channel-estimation-and-decoding via LASSO.

1. INTRODUCTION

When designing a digital communications receiver, it is common to model the effects of multipath propagation in discrete time using a convolutive linear channel that, in the slow-fading scenario, can be characterized by a fixed impulse response $\{x_j\}_{j=0}^{L-1}$ over the duration of one codeword. As the communication bandwidth increases, the channel taps tend to be heavy-tailed or "sparse" in that only a few values in $\{x_j\}_{j=0}^{L-1}$ have significant amplitude [1]. Moreover, groups of large taps are often clustered together in lag j. Furthermore, both sparsity and clustering can be lag-dependent, such as when the receiver's timing-synchronization mechanism aligns the first strong multipath arrivals with a particular reference lag j.

Recently, there have been many attempts to apply breakthrough non-linear estimation techniques from the field of "compressive sensing" (e.g., LASSO [2]) to the wireless channel estimation problem [3]. These works take a *decoupled* approach to the problem of channel estimation and data decoding, in that pilot-symbol knowledge is first exploited for sparse-channel estimation, after which the channel estimate is used for data decoding. However, this decoupled approach is known to be sub-optimal when the taps are non-Gaussian.

The considerations above motivate a *joint* approach to structured-sparse-channel-estimation and decoding that offers both near-optimal decoding performance and low implementation complexity. In this paper, we propose exactly such a scheme. In particular, we focus on orthogonal frequency-division multiplexing (OFDM) with bit-interleaved coded modulation (BICM), and propose a novel factor-graph-based receiver that leverages recent results in "generalized approximate message passing" (GAMP) [4], soft-input/soft-output (SISO) decoding [5], and structured-sparse estimation [6]. Our receiver assumes a clustered-sparse channel-tap prior constructed using a two-state Gaussian mixture with a Markov model on the hidden tap state. The scheme that we propose has only $\mathcal{O}(N \log_2 N + N2^M)$ complexity, where N denotes the number of subcarriers and M denotes the number of bits per subcarrier, facilitating large values of N and channel length $L < N$ (e.g., we use $N = 1024$ and $L = 256$ for our numerical results). For rich non-line-of-sight (NLOS) channels generated according to the 802.15.4a standard [7], our numerical experiments show bit error rate (BER) performance within 1 dB of the known-channel bound and 4 dB better than decoupled channel-estimation-and-decoding via LASSO.

Our work differs from existing factor-graph approaches to joint channel-estimation and decoding (JCED) for OFDM (e.g., [8–11]) in that we use i) a sparse (i.e., non-Gaussian) channel-tap prior, ii) a clustered (i.e., non-independent) channel-tap prior, and iii) the GAMP algorithm, which was proposed and rigorously analyzed as $N, L \to \infty$ in [4]. Although we focus on the case of clustered-sparse channels, our approach could be applied to non-sparse (i.e., Gaussian) or non-clustered (i.e., independent) channel-taps or, e.g., non-sparse channels with unknown length L [8], with minor modifications of our assumed channel prior. Finally, we mention that this work is an evolution of [11], which applied an $\mathcal{O}(NL)$-complexity "relaxed belief propagation" algorithm to unclustered exactly sparse channels, and an abbreviated version of [12].[1]

2. SYSTEM MODEL

2.1. The BICM-OFDM model

We consider an OFDM system with N subcarriers, each modulated by a QAM symbol from a 2^M-ary unit-energy constellation $\mathbb{S}$. Of the N subcarriers, N_{p} are dedicated as pilots, and the remaining $N_{\mathsf{d}} \triangleq N - N_{\mathsf{p}}$ are used to transmit a total of M_{t} training bits and $M_{\mathsf{d}} \triangleq N_{\mathsf{d}} M - M_{\mathsf{t}}$ coded/interleaved data bits. The data bits are generated by encoding M_{i} information bits using a rate-R coder, interleaving them, and partitioning the resulting $M_{\mathsf{c}} \triangleq M_{\mathsf{i}}/R$ bits among an integer number $Q \triangleq M_{\mathsf{c}}/M_{\mathsf{d}}$ of OFDM symbols. We note that the resulting scheme has a spectral efficiency of $\eta \triangleq M_{\mathsf{d}} R/N$ information bits per channel use (bpcu).

This work was supported in part by NSF grant CCF-1018368 and DARPA/ONR grant N66001-10-1-4090, and an allocation of computing time from the Ohio Supercomputer Center.

[1] In the sequel, we use $\mathrm{Re}(\cdot)$ to extract the real part, $\delta(\tau)$ to denote the Dirac impulse, and δ_n to denote the Kronecker impulse. Furthermore, $\mathcal{D}(\boldsymbol{b})$ constructs a diagonal matrix from the vector $\boldsymbol{b}$, and $\mathcal{CN}(x; \hat{x}, \mu^x) \triangleq (\pi\mu^x)^{-1} \exp(-|x - \hat{x}|^2/\mu^x)$ denotes the circular Gaussian pdf with mean $\hat{x}$ and variance μ^x. We often use $(\hat{v}_j, \mu_j^v)$ when referring to the mean and variance of random variable V_j.

In the sequel, we use $s^{(k)} \in \mathbb{S}$ for $k \in \{1,\dots,2^M\}$ to denote the k^{th} element of the QAM constellation, and $\boldsymbol{c}^{(k)} \triangleq [c_1^{(k)},\dots,c_M^{(k)}]^\mathsf{T}$ to denote the corresponding bits as defined by the symbol mapping. Likewise, we use $s_i[q] \in \mathbb{S}$ for the QAM symbol transmitted on the i^{th} subcarrier of the q^{th} OFDM symbol and $\boldsymbol{c}_i[q] \triangleq [c_{i,1}[q],\dots,c_{i,M}[q]]^\mathsf{T}$ for the coded/interleaved bits corresponding to that symbol. We denote the coded/interleaved bits in the q^{th} OFDM symbol by $\boldsymbol{c}[q] \triangleq [\boldsymbol{c}_0[q],\dots,\boldsymbol{c}_{N-1}[q]]^\mathsf{T}$ and those in the entire (interleaved) codeword by $\boldsymbol{c} \triangleq [\boldsymbol{c}[1],\dots,\boldsymbol{c}[Q]]^\mathsf{T}$. The elements of $\boldsymbol{c}$ that are apriori known as pilot or training bits will be referred to as $\boldsymbol{c}_{\mathsf{pt}}$. The remainder of $\boldsymbol{c}$ is determined from the information bits $\boldsymbol{b} \triangleq [b_1,\dots,b_{M_\mathsf{i}}]^\mathsf{T}$ by coding/interleaving.

To modulate the q^{th} OFDM symbol, an N-point inverse discrete Fourier transform (DFT) $\boldsymbol{\Phi}^\mathsf{H}$ is applied to the QAM sequence $\boldsymbol{s}[q] = [s_0[q],\dots,s_{N-1}[q]]^\mathsf{T}$, yielding the time-domain sequence $\boldsymbol{\Phi}^\mathsf{H}\boldsymbol{s}[q] = \boldsymbol{a}[q] = [a_0[q],\dots,a_{N-1}[q]]^\mathsf{T}$. The OFDM waveform $a(t)$ is then constructed using L-cyclic-prefixed versions of $\{a_j[q]\}$ and the transmission pulse $g_\mathsf{t}(\tau)$ using the standard procedure, i.e., $a(t) = \sum_{q=1}^{Q}\sum_{j=-L}^{N-1} a_{\langle j\rangle_N}[q]\, g_\mathsf{t}\big(t - jT - q(N+L)T\big)$, where T denotes the baud interval (in seconds) and $L < N$.

The waveform $a(t)$ propagates through a noisy channel with an impulse response $h(\tau)$ that is supported on the interval $[\tau_{\min},\tau_{\max}]$, resulting in the received waveform $r(t) = v(t) + \int_{\tau_{\min}}^{\tau_{\max}} h(\tau)a(t-\tau)d\tau$, where $v(t)$ is a Gaussian noise process with flat power spectral density N_o. Using the reception pulse $g_\mathsf{r}(\tau)$, The receiver obtains the samples $r_j[q] = \int r(t)\, g_\mathsf{r}\big(jT + q(N+L)T - t\big)dt$, and applies an N-DFT $\boldsymbol{\Phi}$ to the sequence $\boldsymbol{r}[q] = [r_0[q],\dots,r_{N-1}[q]]^\mathsf{T}$, yielding $\boldsymbol{y}[q] = [y_0[q],\dots,y_{N-1}[q]]^\mathsf{T} = \boldsymbol{\Phi}\boldsymbol{r}[q]$ for $q = 1\dots Q$.

For the pulse-shaped channel response $x(\tau) \triangleq (g_\mathsf{r} \star h \star g_\mathsf{t})(\tau)$, it is well known (e.g., [13]) that, when the support of $x(\tau)$ is contained in $[0, LT)$, the observation on the i^{th} subcarrier can be written as

$$y_i[q] \;=\; s_i[q]z_i[q] + v_i[q], \tag{1}$$

where $z_i[q] \in \mathbb{C}$ is the i^{th} subcarrier's gain and $\{v_i[q]\}$ are Gaussian noise samples. Furthermore, defining the uniformly sampled channel "taps" $x_j[q] \triangleq x(jT + q(N+L)T)$, the subcarrier gains are related to these taps through the DFT: $z_i[q] = \sum_{j=0}^{L-1}\Phi_{ij}x_j[q]$. In addition, when $(g_\mathsf{r} \star g_\mathsf{t})(\tau)$ is a Nyquist pulse, $\{v_i[q]\}_{\forall i,q}$ are statistically independent with variance $\mu^v = N_o$.

To simplify the development, we assume that $Q = 1$ in the sequel (but not in the simulations), and drop the index $[q]$ for brevity.

2.2. A clustered-sparse tap prior

For our message-passing-based receiver, we seek a prior on the channel taps $\{x_j\}$ that is capable of representing the lag-dependent clustered sparsity described in Section 1. For this purpose, we use the two-state Gaussian mixture (GM2) prior

$$p(x_j) \;=\; (1-\lambda_j)\mathcal{CN}(x_j; 0, \mu_j^0) + \lambda_j\mathcal{CN}(x_j; 0, \mu_j^1), \tag{2}$$

where $\mu_j^0 \geq 0$ denotes the variance while in the "small" state, $\mu_j^1 > \mu_j^0$ denotes the variance while in the "big" state, and $\lambda_j \triangleq \Pr\{d_j = 1\}$ denotes the prior probability of x_j being in the "big" state. Here, we use $d_j \in \{0,1\}$ to denote the hidden state. For example, if x_j was presumed to be a "sparse" tap, then we would choose $\lambda_j \ll 1$ and $\mu_j^1 \gg \mu_j^0$ in (2). If, on the other hand, x_j is presumed to be (non-sparse) Rayleigh-fading, we would choose $\lambda_j = 1$ and set μ_j^1 equal to the tap variance.

To capture the big-tap clustering behavior, we employ a hidden Markov model (HMM), i.e., we model the tap states $\{d_j\}_{j=0}^{L-1}$ as a Markov chain (MC) with switching probabilities $p_j^{01} \triangleq \Pr\{d_{j+1} = 0 \,|\, d_j = 1\}$ and $p_j^{10} \triangleq \Pr\{d_{j+1} = 1 \,|\, d_j = 0\}$. Here, $p_j^{01} < 0.5$ implies that the neighbors of a big x_j tend to be big, and $p_j^{10} < 0.5$ implies that the neighbors of a small x_j tend to be small. The MC implies that $\lambda_{j+1} = \lambda_j(1-p_j^{01}) + (1-\lambda_j)p_j^{10}$ must hold for all j. Although we allow correlation among the tap states, we assume that the tap *amplitudes* are conditionally independent, i.e., $p(x_{j+1}, x_j \,|\, d_{j+1}, d_j) = p(x_j \,|\, d_j)p(x_{j+1} \,|\, d_{j+1})$. Our experiences with IEEE 802.15.4a channels suggest that this is a valid assumption.

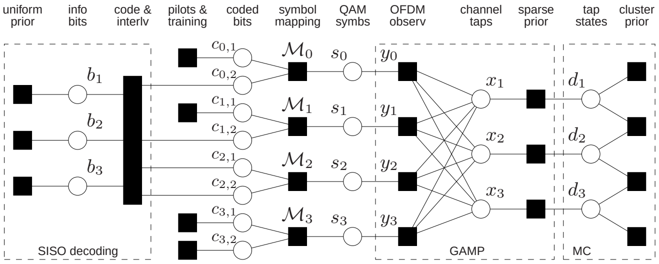

Fig. 4. Factor graph of the JCED problem for a toy example with $M_\mathsf{i} = 3$ information bits, $N_\mathsf{p} = 1$ pilot subcarrier (at subcarrier index $i = 3$), $M_\mathsf{t} = 2$ training bits, $M = 2$ bits per QAM symbol, $N = 4$ OFDM subcarriers, and channel impulse response length $L = 3$.

2.3. An illustrative example: IEEE 802.15.4a channels

As an example of lag-dependent clustered sparsity, we examine tap vectors $\boldsymbol{x} \triangleq [x_0,\dots,x_{L-1}]^\mathsf{T}$ created from square-root raised-cosine (SRRC) pulses $\{g_\mathsf{r}(\tau), g_\mathsf{t}(\tau)\}$ with parameter 0.5, and channel impulse responses $h(\tau)$ generated as specified for the "outdoor non-line-of-sight" case of the 802.15.4a standard [7]. We assumed the bandwidth $T^{-1} = 256$ MHz, for which $L = 256$ taps capture all significant energy in $h(\tau)$, and timing synchronization so that the first significant multipath arrival falls in the vicinity of $j = L_\mathsf{pre} = 21$.

We now describe an experiment conducted using 10000 realizations of the random tap vector $\boldsymbol{x}$. To start, Fig. 1 shows histograms of $\mathrm{Re}(x_j)$ for lags $j \in \{5, 23, 128, 230\}$. There it can be seen that the distribution of $\mathrm{Re}(x_j)$ changes significantly with lag j: for precursor lags $j < L_\mathsf{pre}$, it looks Gaussian; for near-cursor lags $j \approx L_\mathsf{pre}$, it looks Laplacian; and, for post-cursor lags $j \gg L_\mathsf{pre}$, it looks extremely heavy-tailed. In Fig. 2, we plot a typical realization of $\boldsymbol{x}$ and notice clustering among the big taps. For comparison, we also plot the empirical power-delay profile (PDP) $\boldsymbol{\rho} \triangleq [\rho_0,\dots,\rho_{L-1}]^\mathsf{T}$, where $\rho_j \triangleq \mathrm{E}\{|x_j|^2\}$. Using the EM algorithm to fit the GM2 parameters $\{\lambda_j, \mu_j^0, \mu_j^1\}_{j=0}^{L-1}$ [14, p. 435], we get the big-variance profile $\boldsymbol{\mu}^1 \triangleq [\mu_0^1,\dots,\mu_{L-1}^1]^\mathsf{T}$ and small-variance profile $\boldsymbol{\mu}^0$ shown in Fig. 2, and the sparsity profile $\boldsymbol{\lambda} \triangleq [\lambda_0,\dots,\lambda_{L-1}]^\mathsf{T}$ shown in Fig. 3. We observe that the resulting GM2 parameters show a peak in λ_j near $j = L_\mathsf{pre}$ and sparsity increasing with j. The switching probabilities $\boldsymbol{p}^{01} \triangleq [p_0^{01},\dots,p_{L-1}^{01}]^\mathsf{T}$ and $\boldsymbol{p}^{10}$ shown in Fig. 3 were then estimated from realizations of the MAP-detected state vector $\boldsymbol{d} \triangleq [d_0,\dots,d_{L-1}]^\mathsf{T}$, where Fig. 2 shows the MAP detection threshold. Finally, normalized estimates of the conditional correlation $\mathrm{E}\{x_{j+1}x_j^* \,|\, d_{j+1} = 1, d_j = 1\}$ were found to have magnitude < 0.1, validating our conditional-independence assumption.

3. JOINT CHANNEL ESTIMATION AND DECODING

Our goal is to infer the information bits $\boldsymbol{b}$ from the OFDM observations $\boldsymbol{y}$ and the pilot/training bits $\boldsymbol{c}_\mathsf{pt}$, without knowing the channel state $\boldsymbol{x}$. In particular, we aim to maximize the posterior pmf

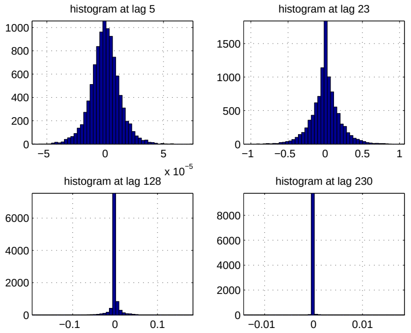

Fig. 1. Histograms of $\mathrm{Re}(x_j)$ for lags $j \in \{5, 23, 128, 230\}$, with "tight" axes.

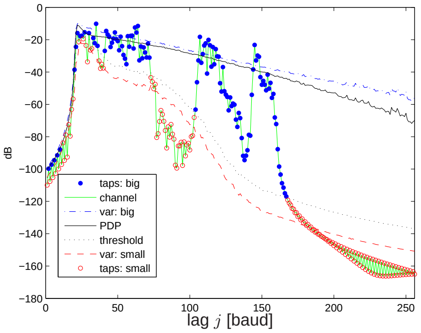

Fig. 2. A realization of $\{x_j\}$ generated from the IEEE 802.15.4a model with SRRC pulse shaping.

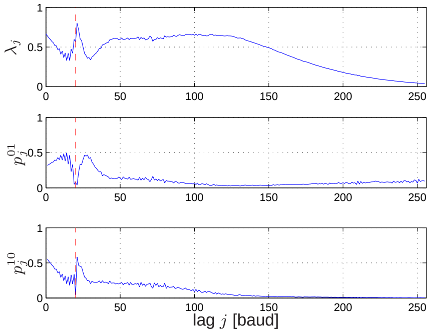

Fig. 3. Empirically estimated statistics on the tap-states $\{d_j\}$. Top: λ_j, middle: p_j^{01}, bottom: p_j^{10}.

$p(b_m \mid \boldsymbol{y}, \boldsymbol{c}_{\text{pt}})$ of each info bit. To exploit prior knowledge that $\boldsymbol{x}$ is clustered-sparse, we employ the GM2-HMM prior described in Section 2.2, yielding the *factor graph* in Fig. 4, where the round nodes represent random variables and the square nodes represent the probabilistic relationships between them.

3.1. Background on belief propagation and GAMP

Although exact evaluation of the posteriors $\{p(b_m \mid \boldsymbol{y}, \boldsymbol{c}_{\text{pt}})\}$ is computationally impractical for the problem sizes of interest, these posteriors can be approximately evaluated using *belief propagation* (BP) on the loopy factor graph in Fig. 4. In textbook BP, beliefs take the form of pdfs/pmfs that are propagated among nodes of the factor graph via the *sum/product algorithm* (SPA) [14]. When the factor graph contains no loops, SPA-BP yields exact posteriors after two rounds of message passing (i.e., forward and backward). But, in the presence of loops, convergence to the exact posteriors is not guaranteed. Even so, there exist many problems to which loopy BP has been successfully applied, including LDPC decoding [5], Markov field inference [14], and compressed sensing [4, 6, 15].

An important sub-problem within our larger bit-inference problem is the estimation of a vector of independent non-Gaussian variables $\boldsymbol{x}$ that are linearly mixed via $\boldsymbol{\Phi} \in \mathbb{C}^{N \times L}$ to form $\boldsymbol{z} = \boldsymbol{\Phi x} = [z_0, \ldots, z_{N-1}]^{\mathsf{T}}$, and subsequently observed as noisy measurements $\boldsymbol{y}$ through possibly non-Gaussian pdfs $\{p_{Y_i|Z_i}(.\,|\,.)\}$. In our case, (2) specifies the non-Gaussian prior on x_j and (1)—given the finite-alphabet uncertainty in s_i—yields the non-Gaussian measurement pdf $p_{Y_i|Z_i}$. This "linear mixing" sub-problem is described by the factor graph shown within the middle dashed box in Fig. 4, where each node "y_i" represents the measurement pdf $p_{Y_i|Z_i}$ and the node to the right of each node "x_j" represents the GM2 prior on x_j. Building on recent work on message passing algorithms for compressed sensing [15], Rangan recently proposed a so-called *generalized approximate message passing* (GAMP) scheme that, for the sub-problem described above, admits rigorous analysis as $N, L \to \infty$ [4]. The GAMP algorithm[2] is summarized in Table 1.

3.2. Joint estimation and decoding using GAMP

We now detail our application of GAMP to joint channel-estimation and decoding (JCED) under the GM2-HMM tap prior, frequently referring to the factor graph in Fig. 4.

Because our factor graph is loopy, there exists considerable freedom in the message passing schedule. Roughly speaking, we choose to pass messages from the left to the right of Fig. 4 and back again, several times, stopping as soon as the messages converge. Each of these full cycles of message passing will be referred to as a "turbo iteration." However, during a single turbo iteration, there may be multiple iterations of message passing *between* the GAMP and MC sub-graphs, which will be referred to as "equalizer" iterations. Furthermore, during a single equalizer iteration, there may be multiple iterations of message passing *within* the GAMP sub-graph, while there is at most one forward-backward iteration *within* the MC sub-graph. The message passing details are discussed below.

```
definitions:
    $p_{Z_i|Y_i}(z|y;\hat{z},\mu^z) = \frac{p_{Y_i|Z_i}(y|z)\,\mathcal{CN}(z;\hat{z},\mu^z)}{\int_{z'} p_{Y_i|Z_i}(y|z')\,\mathcal{CN}(z';\hat{z},\mu^z)}$   (D1)
    $g_{\text{out},i}(y,\hat{z},\mu^z) = \frac{1}{\mu^z}\left(\mathrm{E}_{Z_i|Y_i}\{z|y;\hat{z},\mu^z\} - \hat{z}\right)$   (D2)
    $g'_{\text{out},i}(y,\hat{z},\mu^z) = \frac{1}{\mu^z}\left(\frac{\mathrm{var}_{Z_i|Y_i}\{z|y;\hat{z},\mu^z\}}{\mu^z} - 1\right)$   (D3)
    $p_{X_j}(x;\hat{r},\mu^r) = \frac{p_{X_j}(x)\,\mathcal{CN}(x;\hat{r},\mu^r)}{\int_{x'} p_{X_j}(x')\,\mathcal{CN}(x';\hat{r},\mu^r)}$   (D4)
    $g_{\text{in},j}(\hat{r},\mu^r) = \int_x x\, p_{X_j}(x;\hat{r},\mu^r)$   (D5)
    $g'_{\text{in},j}(\hat{r},\mu^r) = \frac{1}{\mu^r}\int_x |x - g_{\text{in},j}(\hat{r},\mu^r)|^2\, p_{X_j}(x;\hat{r},\mu^r)$   (D6)
initialize:
    $\forall j : \hat{x}_j(1) = \int_x x\, p_{X_j}(x)$   (I1)
    $\forall j : \mu^x_j(1) = \int_x |x - \hat{x}_j(1)|^2 p_{X_j}(x)$   (I2)
    $\forall i : \hat{u}_i(0) = 0$   (I3)
for $n = 1, 2, 3, \ldots$
    $\forall i : \hat{z}_i(n) = \sum_{j=0}^{L-1} \Phi_{ij}\hat{x}_j(n)$   (R1)
    $\forall i : \mu^z_i(n) = \sum_{j=0}^{L-1} |\Phi_{ij}|^2 \mu^x_j(n)$   (R2)
    $\forall i : \hat{p}_i(n) = \hat{z}_i(n) - \mu^z_i(n)\,\hat{u}_i(n-1)$   (R3)
    $\forall i : \hat{u}_i(n) = g_{\text{out},i}(y_i, \hat{p}_i(n), \mu^z_i(n))$   (R4)
    $\forall i : \mu^u_i(n) = -g'_{\text{out},i}(y_i, \hat{p}_i(n), \mu^z_i(n))$   (R5)
    $\forall j : \mu^r_j(n) = \left(\sum_{i=0}^{N-1} |\Phi_{ij}|^2 \mu^u_i(n)\right)^{-1}$   (R6)
    $\forall j : \hat{r}_j(n) = \hat{x}_j(n) + \mu^r_j(n)\sum_{i=0}^{N-1} \Phi^*_{ij}\hat{u}_i(n)$   (R7)
    $\forall j : \mu^x_j(n+1) = \mu^r_j(n) g'_{\text{in},j}(\hat{r}_j(n), \mu^r_j(n))$   (R8)
    $\forall j : \hat{x}_j(n+1) = g_{\text{in},j}(\hat{r}_j(n), \mu^r_j(n))$   (R9)
end
```

Table 1. The GAMP Algorithm

At the start of the first turbo iteration, there is total uncertainty about the information bits, so that $\Pr\{b_m = 1\} = \frac{1}{2}$. Thus, the initial bit beliefs flowing rightward out of the coding/interleaving block are uniformly distributed. Meanwhile, the pilot/training bits are known with certainty, and thus take $\Pr\{b_m = 1\} \in \{0, 1\}$.

Coded-bit beliefs are then propagated rightward into the symbol mapping nodes. Since the symbol mapping is deterministic, the corresponding pdf factors take the form $p(s^{(k)} \,|\, \boldsymbol{c}^{(l)}) = \delta_{k-l}$. The SPA then dictates that the message passed rightward from symbol mapping node "$\mathcal{M}_i$" is $p_{\mathcal{M}_i \to s_i}(s^{(k)}) = \prod_{m=1}^{M} p_{c_{i,m} \to \mathcal{M}_i}(c_m^{(k)})$,

[2] To be precise, the GAMP algorithm in Table 1 is an extension of that proposed in [4]. Table 1 handles circular *complex-valued* distributions and *non-identically* distributed signals and measurements.

which is then copied forward as the message 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 Section 3.1, that the symbol-belief passed rightward into the measurement node "y_i" determines the pdf $p_{Y_i|Z_i}$ used in GAMP. Writing this symbol belief as $\boldsymbol{\beta}_i \triangleq [\beta_i^{(1)}, \ldots, \beta_i^{(2^M)}]^{\mathsf{T}}$ for $\beta_i^{(k)} \triangleq p_{s_i \to y_i}(s^{(k)})$, equation (1) implies the measurement pdf $p_{Y_i|Z_i}(y|z) = \sum_{k=1}^{2^M} \beta_i^{(k)}\, \mathcal{CN}(y; s^{(k)}z; \mu^v)$. Then, defining

$$\xi_i^{(k)}(y, \hat{z}, \mu^z) \triangleq \frac{\beta_i^{(k)} \mathcal{CN}(y; s^{(k)}\hat{z}, |s^{(k)}|^2\mu^z + \mu^v)}{\sum_{k'} \beta_i^{(k')} \mathcal{CN}(y; s^{(k')}\hat{z}, |s^{(k')}|^2\mu^z + \mu^v)} \qquad (3)$$

$$\zeta^{(k)}(\mu^z) \triangleq \frac{|s^{(k)}|^2\mu^z}{|s^{(k)}|^2\mu^z + \mu^v} \qquad (4)$$

$$\hat{e}^{(k)}(y, \hat{z}, \mu^z) \triangleq \big(\tfrac{y}{s^{(k)}} - \hat{z}\big)\zeta^{(k)}(\mu^z) \qquad (5)$$

$$\hat{e}_i(y, \hat{z}, \mu^z) \triangleq \sum_{k=1}^{2^M} \xi_i^{(k)}(y, \hat{z}, \mu^z)\, \hat{e}^{(k)}(y, \hat{z}, \mu^z) \qquad (6)$$

$$\mu_i^e(y, \hat{z}, \mu^z) \triangleq \sum_{k=1}^{2^M} \xi_i^{(k)}(y, \hat{z}, \mu^z)\, \big(|\hat{e}^{(k)}(y, \hat{z}, \mu^z) - \hat{e}_i(y, \hat{z}, \mu^z)|^2 + \tfrac{\mu^v \zeta^{(k)}(\mu^z)}{s^{(k)}}\big), \qquad (7)$$

where $\boldsymbol{\xi}_i \triangleq [\xi_i^{(1)}, \ldots, \xi_i^{(2^M)}]^{\mathsf{T}}$ characterizes the posterior pmf on s_i under the channel model $z_i \sim \mathcal{CN}(\hat{z}, \mu^z)$, it can be shown that the quantities in (D2)-(D3) of Table 1 become

$$g_{\mathsf{out},i}(y, \hat{z}, \mu^z) = \tfrac{1}{\mu^z} \hat{e}_i(y, \hat{z}, \mu^z) \qquad (8)$$

$$g'_{\mathsf{out},i}(y, \hat{z}, \mu^z) = \tfrac{1}{\mu^z}\big(\tfrac{\mu_i^e(y, \hat{z}, \mu^z)}{\mu^z} - 1\big). \qquad (9)$$

Likewise, defining $\gamma_j^0(\mu^r) \triangleq (1 + \mu^r/\mu_j^0)^{-1}$ and $\gamma_j^1(\mu^r) \triangleq (1 + \mu^r/\mu_j^1)^{-1}$ and $\alpha_j(\hat{r}, \mu^r) \triangleq 1/\big(1 + (\mathcal{L}_j^{\mathsf{apri}} \cdot \mathcal{L}_j^{\mathsf{ext}}(\hat{r}, \mu^r))^{-1}\big)$, where $\mathcal{L}_j^{\mathsf{apri}} \triangleq \frac{\lambda_j}{1-\lambda_j}$ is the apriori likelihood ratio $\frac{\Pr\{d_j=1\}}{\Pr\{d_j=0\}}$ on the hidden state, $\mathcal{L}_j^{\mathsf{ext}}(\hat{r}, \mu^r) \triangleq \frac{\mathcal{CN}(\hat{r}; 0, \mu_j^1 + \mu^r)}{\mathcal{CN}(\hat{r}; 0, \mu_j^0 + \mu^r)}$ is GAMP's extrinsic likelihood ratio, and $\alpha_j(\hat{r}, \mu^r)$ is the posterior $\Pr\{d_j = 1\}$, it can be shown from (2) that the quantities in (D5)-(D6) of Table 1 become

$$g_{\mathsf{in},j}(\hat{r}, \mu^r) = \big(\alpha_j\, \gamma_j^1 + (1 - \alpha_j)\, \gamma_j^0\big)\hat{r} \qquad (10)$$

$$g'_{\mathsf{in},j}(\hat{r}, \mu^r) = \alpha_j(1 - \alpha_j)(\gamma_j^1 - \gamma_j^0)^2\, |\hat{r}|^2/\mu^r + \alpha_j\gamma_j^1 + (1 - \alpha_j)\gamma_j^0, \qquad (11)$$

Using (3)-(11), the GAMP algorithm in Table 1 is iterated until it converges. In doing so, GAMP generates (a close approximation to) both the conditional means $\hat{\boldsymbol{x}}$ and variances $\boldsymbol{\mu}^x \triangleq [\mu_0^x, \ldots, \mu_{L-1}^x]^{\mathsf{T}}$ given the observations $\boldsymbol{y}$, the soft symbol priors $\boldsymbol{\beta} \triangleq [\boldsymbol{\beta}_0, \ldots, \boldsymbol{\beta}_{L-1}]^{\mathsf{T}}$ and the sparsity prior $\boldsymbol{\lambda}$. Conveniently, GAMP also returns both the conditional means $\hat{\boldsymbol{z}}$ and variances $\boldsymbol{\mu}^z$ of the subchannel gains $\boldsymbol{z}$.

It should be noted that, due to unit-modulus property of the DFT elements Φ_{ij}, step (R2) in Table 1 simplifies to $\mu_i^z(n) = \sum_j \mu_j^x(n)$ and (R6) simplifies to $\mu_j^r(n) = (\sum_i \mu_i^u(n))^{-1}$. The complexity of GAMP is then dominated by either the matrix-vector products $\sum_j \Phi_{ij}\hat{x}_j(n)$ in (R1) and $\sum_i \Phi_{ij}^* \hat{u}_i(n)$ in (R7), which can be implemented using a $N \log_2 N$-multiply FFT when N is a power-of-two, or by the calculation of $\{\hat{e}_i, \mu_i^e\}_{i=0}^{N-1}$ in (6)-(7), which requires $\mathcal{O}(N2^M)$ multiplies. Thus, GAMP requires only $\mathcal{O}(N \log_2 N + N2^M)$ multiplies per iteration.

After the messages within the GAMP sub-graph have converged, tap-state beliefs are passed rightward to the MC sub-graph. In particular, the SPA dictates that GAMP pass the extrinsic likelihood ratios $\mathcal{L}_j^{\mathsf{ext}}$. Since the MC sub-graph is non-loopy, only one iteration of forward-backward message passing (see [14]) is sufficient, after which the resulting tap-state likelihoods are passed leftward back to GAMP, where they are treated as tap-state priors $\boldsymbol{\lambda}$ in the next equalizer iteration. This interaction between the GAMP and MC blocks is essentially the structured-sparse reconstruction scheme in [6].

When the tap-state likelihoods passed between GAMP and MC have converged, the equalizer iterations are terminated and messages are passed leftward from the GAMP block. For this, SPA dictates

$$p_{s_i \leftarrow y_i}(s) = \mathcal{CN}(y_i; s\hat{z}_i, |s|^2\mu_i^z + \mu^v), \qquad (12)$$

where $(\hat{z}_i, \mu_i^z)$ play the role of soft channel estimates. Furthermore, the SPA 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 $c_{i,m}$. They take the form $p_{c_{i,m} \leftarrow \mathcal{M}_i}(c) = \frac{1}{p_{c_{i,m} \to \mathcal{M}_i}(c)} \sum_{k: c_m^{(k)} = c} p_{\mathcal{M}_i \leftarrow s_i}(s^{(k)}) p_{\mathcal{M}_i \to s_i}(s^{(k)})$ for pairs (i, m) that do not correspond to pilot/training bits.

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) decoder/deinterleaver, which treats them as priors. Since SISO decoding is a well-studied topic [5], we will not give 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 coding/interleaving block to begin another turbo iteration. These turbo iterations 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

We now present numerical results comparing our GAMP-based JCED scheme to decoupled channel-estimation and decoding (DCED) based on pilot-aided linear MMSE (LMMSE) channel estimates, LASSO channel estimates [2], and perfect channel state information (CSI).

Setup: For all results, we used irregular LDPC codes with codeword length ~ 10000 and average column weight 3, generated (and decoded) using the publicly available software [16], with random interleaving. We focus on the case of $N = 1024$ subcarriers and 16-QAM (i.e., $M = 4$) operating at a spectral efficiency of $\eta = 2$ bpcu. For bit-to-symbol mapping, we used multilevel Gray-mapping. In some simulations, we used $N_{\mathsf{p}} > 0$ pilot-only subcarriers and $M_{\mathsf{t}} = 0$ interspersed training bits, whereas in others we used $N_{\mathsf{p}} = 0$ and $M_{\mathsf{t}} > 0$. When $N_{\mathsf{p}} > 0$, the pilot subcarriers were placed randomly and modulated with (known) QAM symbols chosen uniformly at random. When $M_{\mathsf{t}} > 0$, unit-valued training bits were placed at the most significant bits (MSBs) of uniformly spaced data-subcarriers.

Realizations of the tap vector $\boldsymbol{x}[q]$ were generated from 802.15.4a outdoor-NLOS impulse responses and SRRC pulses, as described in Section 2.3. All reported results are averaged over ≥ 1000 channel realizations (i.e., $\gtrsim 10^7$ info bits).

The GM2-HMM prior parameters used by GAMP were fit from 10000 realizations of the tap-vector $\boldsymbol{x}$ using the procedure described in Section 2.3. For JCED-GAMP, we used a *maximum* of 20 turbo iterations, 5 equalizer iterations, 15 GAMP iterations, and 25 LDPC decoder iterations, although these maxima were seldom reached.

The following procedure was used to implement DCED from a given set of tap estimates $\{\hat{\boldsymbol{x}}[q]\}_{q=1}^{Q}$. First, the subcarrier estimates $\hat{\boldsymbol{z}}[q] = \boldsymbol{\Phi}\hat{\boldsymbol{x}}[q]$ were computed, from which the (genie-aided empirical) variance $\hat{\mu}^z[q] \triangleq \|\hat{\boldsymbol{z}}[q] - \boldsymbol{z}[q]\|_2^2/N$ was calculated. Then, using the soft channel estimates $\{\hat{\boldsymbol{z}}[q], \hat{\mu}^z[q]\}_{q=1}^{Q}$, leftward SPA-BP on the factor graph in Fig. 4 was performed exactly as described in Section 3.2. LMMSE tap estimates were computed via $\hat{\boldsymbol{x}}_{\mathsf{lmmse}}[q] = \boldsymbol{A}^{\mathsf{H}}[q]\big(\boldsymbol{A}[q]\,\mathcal{D}(\boldsymbol{\rho})\boldsymbol{A}^{\mathsf{H}}[q] + \mu^v\boldsymbol{I}\big)^{-1}\boldsymbol{y}_{\mathsf{pt}}[q]$, where $\boldsymbol{s}_{\mathsf{pt}}[q] \in \mathbb{S}^{N_{\mathsf{p}}}$ are the

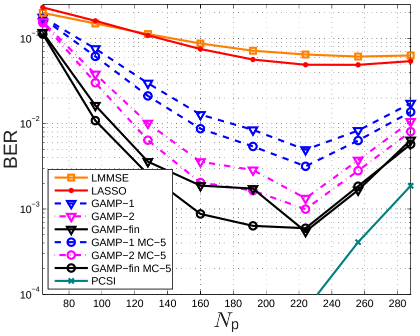

Fig. 5. BER versus number of pilot subcarriers N_{p}, for $E_b/N_o = 11$ dB, $M_{\text{t}} = 0$ training bits, $\eta = 2$ bpcu, and 16-QAM.

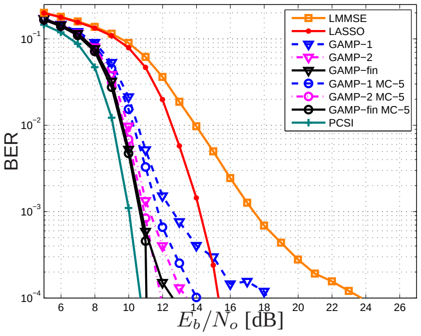

Fig. 6. BER versus E_b/N_o, for $N_{\text{p}} = 224$ pilot subcarriers, $M_{\text{t}} = 0$ training bits, $\eta = 2$ bpcu, and 16-QAM.

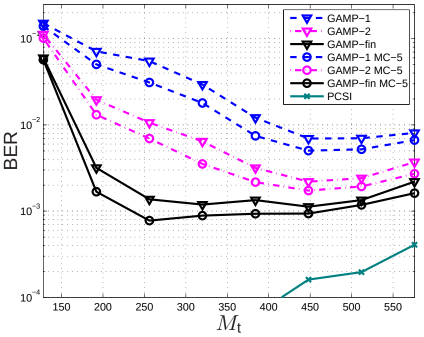

Fig. 7. BER versus number of interspersed training bits M_{t}, for $E_b/N_o = 10$ dB, $N_{\text{p}} = 0$ pilots subcarriers, $\eta = 2$ bpcu, and 16-QAM.

pilot symbols, $\boldsymbol{y}_{\text{pt}}[q] \in \mathbb{C}^{N_{\text{p}}}$ are the pilot subcarrier observations, $\boldsymbol{\rho}$ is the PDP, $\boldsymbol{A}[q] \triangleq \mathcal{D}(\boldsymbol{s}_{\text{pt}}[q])\boldsymbol{\Phi}_{\text{pt}}$, and $\boldsymbol{\Phi}_{\text{pt}} \in \mathbb{C}^{N_{\text{p}} \times L}$ is constructed from the pilot rows and the first L columns of the N-DFT matrix. LASSO tap estimates $\hat{\boldsymbol{x}}_{\text{lasso}}[q]$ were computed by the celebrated SPGL1 algorithm [17] using $\boldsymbol{A}[q]$ and a genie-optimized tuning parameter.

Figure 5 shows bit error rate (BER) versus the number of pilot subcarriers N_{p} at $E_b/N_o = 11$ dB and a fixed spectral efficiency of $\eta = 2$ bpcu. In this and other figures, "GAMP-# MC-5," refers to JCED-GAMP with # turbo iterations and 5 equalizer iterations, whereas "GAMP-#" alone indicates that the MC block was disconnected (i.e., there was no attempt to exploit tap clustering).

The curves in Fig. 5 exhibit a "notched" shape because, as N_{p} increases, the code rate R must decrease to maintain the fixed spectral efficiency $\eta = 2$ bpcu; while an increase in N_{p} generally makes channel estimation easier, the reduction in R makes data decoding more difficult. For all schemes under comparison, Fig. 5 suggests that the choice $N_{\text{p}} \approx 224$ is optimal under the chosen operating conditions. Overall, we see GAMP significantly outperforming both LMMSE and LASSO even after one turbo iteration. Moreover, we see a noticeable gain from the use of the MC block during the first two turbo iterations, and after turbo convergence if too few pilots are used (i.e., $N_{\text{p}} < 224$).

Figure 6 shows BER versus E_b/N_o using $N_{\text{p}} = 224$ pilot subcarriers (as suggested by Fig. 5) and $M_{\text{t}} = 0$. There, we see LASSO performing about 5 dB from perfect-CSI, and LMMSE performing significantly worse. Remarkably, we see GAMP performing within 1 dB of the perfect-CSI bound (and within 1.5 dB after only 2 turbo iterations). The proximity of the perfect-CSI and GAMP BER traces confirms that the proposed GM2-HMM prior does an excellent job of capturing the lag-dependent clustered-sparse characteristics of the true channel taps. Consistent with Fig. 5, we see that GAMP benefits significantly from the use of the MC block during the initial turbo iterations, and less significantly after turbo convergence.

Although $N_{\text{p}} > 0$ pilot subcarriers are required for DCED channel estimation, JCED can function with $N_{\text{p}} = 0$ as long as $M_{\text{t}} > 0$ interspersed training bits are used. To examine the latter case, Fig. 7 shows BER versus M_{t} at $E_b/N_o = 10$ dB, a fixed spectral efficiency of $\eta = 2$ bpcu, and $N_{\text{p}} = 0$. There we see that there is a relatively wide tolerance on M_{t}, although the value $M_{\text{t}} \approx 450$ appears best when convergence speed is taken into account. Moreover, we can see a small but noticeable BER improvement when the MC block is used. More importantly, by comparing the BER performance in Fig. 7 to that in Fig. 6 at $E_b/N_o = 10$ dB, we see that the BER is about 6 times lower in Fig. 7. Thus, we conclude that placing training bits at MSBs is more efficient than placing them at dedicated pilot subcarriers.

5. REFERENCES

[1] A. F. Molisch, "Ultrawideband propagation channels—Theory, measurement, and modeling," *IEEE Trans. Veh. Tech.*, vol. 54, pp. 1528–1545, Sep. 2005.

[2] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc. B*, vol. 58, no. 1, pp. 267 – 288, 1996.

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

[4] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," *arXiv:1010.5141*, Oct. 2010.

[5] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*. New York: Cambridge University Press, 2003.

[6] P. Schniter, "Turbo reconstruction of structured sparse signals," in *Proc. Conf. Inform. Science & Syst.*, (Princeton, NJ), Mar. 2010.

[7] A. F. Molisch, K. Balakrishnan, C.-C. Chong, S. Emami, A. Fort, J. Karedal, J. Kunisch, H. Schantz, U. Schuster, and K. Siwiak, "IEEE 802.15.4a channel model—Final report," tech. rep., Document IEEE 802.1504-0062-02-004a, 2005.

[8] C. Novak, G. Matz, and 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), pp. 2433–2436, Apr. 2009.

[9] C. Kneivel, Z. Shi, P. A. Hoeher, and G. Auer, "2D graph-based soft channel estimation for MIMO-OFDM," in *Proc. IEEE Int. Conf. Commun.*, (Cape Town), pp. 1–5, Jul. 2010.

[10] G. E. Kirkelund, C. N. Manchón, L. P. B. Christensen, E. Riegler, and 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.

[11] P. Schniter, "Joint estimation and decoding for sparse channels via relaxed belief propagation," in *Proc. Asilomar Conf. Signals Syst. Comput.*, (Pacific Grove, CA), Nov. 2010.

[12] P. Schniter, "A message-passing receiver for BICM-OFDM over unknown clustered-sparse channels," *arXiv:1101.4724*, Jan. 2011.

[13] L. J. Cimini, Jr., "Analysis and simulation of a digital mobile radio channel using orthogonal frequency division multiplexing," *IEEE Trans. Commun.*, vol. 33, pp. 665–765, July 1985.

[14] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2007.

[15] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *arXiv:1001.3448*, Jan. 2010.

[16] I. Kozintsev, "Matlab programs for encoding and decoding of LDPC codes over GF(2^m)." http://www.kozintsev.net/soft.html.

[17] E. van den Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM J. Scientific Comput.*, vol. 31, no. 2, pp. 890–912, 2008.