A Factor Graph Approach to Joint OFDM Channel Estimation and Decoding in Impulsive Noise Environments

Marcel Nassar, *Member, IEEE*, Philip Schniter, *Fellow, IEEE*, and Brian L. Evans, *Fellow, IEEE*

***Abstract*—We propose a novel receiver for orthogonal frequency division multiplexing (OFDM) transmissions in impulsive noise environments. Impulsive noise arises in many modern wireless and wireline communication systems, such as Wi-Fi and powerline communications, due to uncoordinated interference that is much stronger than thermal noise. We first show that the bit-error-rate optimal receiver jointly estimates the propagation channel coefficients, the noise impulses, the finite-alphabet symbols, and the unknown bits. We then propose a near-optimal yet computationally tractable approach to this joint estimation problem using loopy belief propagation. In particular, we merge the recently proposed "generalized approximate message passing" (GAMP) algorithm with the forward-backward algorithm and soft-input soft-output decoding using a "turbo" approach. Numerical results indicate that the proposed receiver drastically outperforms existing receivers under impulsive noise and comes within 1 dB of the matched-filter bound. Meanwhile, with N tones, the proposed factor-graph-based receiver has only $O(N \log N)$ complexity, and it can be parallelized.**

***Index Terms*—OFDM, sum-product algorithm, generalized approximate message passing (GAMP), iterative receivers, factor-graphs, impulsive noise, uncoordinated interference.**

I. Introduction

THE main impairments to a communication system, whether wireless or wireline, are due to multipath propagation through a physical medium and additive noise. Multipath propagation is commonly modeled as a linear convolution that, in the slow-fading scenario, can be characterized by a channel impulse response $\{h_j\}_{j=0}^{L-1}$ that is fixed over the duration of one codeword. In the well-known "uncorrelated Rayleigh/Ricean-fading" scenario, the (complex-baseband) channel "taps" h_j are modeled as independent circular Gaussian random variables. Similarly, in the equally well-known "additive white Gaussian noise" (AWGN) scenario, the time-domain additive noise samples $\{n_t\}_{\forall t}$ are modeled as independent circular Gaussian random variables [1].

A. Motivation

In this work, we focus on applications where the uncorrelated-Rayleigh/Ricean-fading assumption holds but the AWGN assumption does not. Our work is motivated by extensive measurement campaigns of terrestrial wireless installations wherein the additive noise is *impulsive*, with peak noise amplitudes reaching up to 40 dB above the thermal background noise level [2]–[7]. The noise affecting powerline communications (PLC) has also been shown to be highly impulsive, as well as bursty [8]–[10].

We restrict our attention to systems employing (coded or uncoded) orthogonal frequency division multiplexing (OFDM) [1], as used in many modern cellular wireless standards (e.g., IEEE802.11n and LTE) and PLC standards (e.g., PRIME and IEEE1901). OFDM is advantageous in that it facilitates data communication across convolutive multipath channels with high spectral efficiency and low complexity.

The impulsivity of noise has particular consequences for OFDM systems. Recall that, in conventional OFDM receivers, the time-domain received signal is converted to the frequency domain through a discrete Fourier transform (DFT) [1], after which each subcarrier (or "tone") is demodulated independently. Such tone-by-tone demodulation is in fact optimal with AWGN and perfect channel estimates [1], and is highly desirable from a complexity standpoint, since it leaves the DFT as the primary source of receiver complexity, and thus requires only $O(N \log N)$ multiplies per symbol for N tones. When the time-domain noise is impulsive, however, the corresponding frequency-domain noise samples will be highly dependent, and tone-by-tone demodulation is no longer optimal. We are thus strongly motivated to find near-optimal demodulation strategies that preserve the $O(N \log N)$ complexity of classical OFDM. In this work, we propose one such solution that exploits recent breakthroughs in loopy belief propagation.

B. Prior Work

1) OFDM Reception in Impulsive Noise: One popular approach to OFDM reception in impulsive noise stems from

Manuscript received June 07, 2013; revised October 15, 2013; accepted November 27, 2013. Date of publication December 13, 2013; date of current version February 26, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Ignacio Santamaria. The work of P. Schniter was supported in part by NSF grant CCF-1218754, NSF grant CCF-1018368, and by DARPA/ONR grant N66001-10-1-4090. The work of M. Nassar and B. L. Evans was supported by a grant funding from the Semiconductor Research Corporation under SRC GRC Task 1836.063 with liaisons Freescale Semiconductor, IBM, and Texas Instruments. This work was done while M. Nassar was at UT Austin.

M. Nassar is with the Mobile Solutions Lab, Samsung Research America, San Diego, CA 92037 USA (e-mail: mnassar@utexas.edu).

P. Schniter is with the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210 USA (e-mail: schniter@ece.osu.edu).

B. L. Evans is with the Wireless Networking and Communications Group, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: bevans@ece.utexas.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2013.2295063

the argument that the noiseless time-domain received OFDM samples can be modeled as i.i.d Gaussian (according to the central limit theorem with sufficiently many tones), in which case the noise impulses can be detected using a simple threshold test. This approach straightforwardly leads to a decoupled procedure for impulse mitigation and OFDM reception: the time-domain received signal is *pre-processed* via clipping or blanking techniques [11], [12] or (nonlinear) MMSE estimation [13], and the result passed to a conventional DFT receiver for decoding. While agreeable from a complexity standpoint, these techniques give relatively poor communication performance, especially when the power of the impulsive noise is comparable to the power of the OFDM signal, or when higher order modulations are used [13]. This loss of performance can be explained by the fact that the OFDM signal structure is not exploited for noise mitigation. In an attempt to improve communication performance, it has been suggested to iterate between such pre-processing and OFDM decoding, but the approaches suggested to date (e.g., [14]–[17]) have shown limited success, mainly because the adaptation of preprocessing with each iteration is challenging and often done in an ad-hoc manner.

Another popular approach models the time-domain impulsive noise sequence as a sparse vector and then uses *sparse-reconstruction* techniques to estimate this sequence from the observed OFDM null and pilot (i.e., known) tones. The recovered impulse vector is then subtracted from the time-domain received signal, and the result is passed to a conventional DFT receiver for decoding. Algebraic techniques were proposed in [18]–[20], and sparse reconstruction techniques based on compressive-sensing were proposed in [21], [22]. With typical numbers of known tones, these techniques have been shown to work well for very sparse impulsive noise sequences (e.g., one impulse in a 256-tone OFDM system with 30 known tones) but not for practical sparsity rates [21], [23]. Recently, under the highly restrictive condition requiring no overlap between the support of impulsive noise and channel impulse response, this approach was extended to incorporate channel estimation [24].

A more robust approach was proposed in [23], which performs joint symbol detection and impulse-noise estimation using sparse Bayesian learning (SBL). Because [23] decouples channel estimation from impulse-noise estimation and symbol detection, and because it integrates coding in an ad-hoc manner, there is considerable room for improvement. In addition, it performs matrix inversion that is impractical for typical OFDM receivers with hundreds of tones.

2) Factor Graph Receivers: Factor-graph-based receivers [25] have been proposed as a computationally efficient means of tackling the difficult task of *joint* channel, symbol, and bit (JCSB) estimation. Here, messages (generally in the form of pdfs) are passed among the nodes of the factor graph according to belief propagation strategies like the sum-product algorithm (SPA) [26]. Due to the loopy nature of the OFDM factor graph, however, exact implementation of the sum product algorithm is infeasible, and so various approximations have been proposed [27]–[30]. Notably, [30] merged the "generalized approximate message passing" (GAMP) algorithm [31] with a soft-input soft-output decoder in a "turbo" configuration to accomplish near-optimal[1] joint structured-sparse-channel estimation and decoding of bit-interleaved coded-modulation (BICM)-OFDM with $O(N \log N)$ complexity. To our knowledge, no factor-graph-based OFDM receivers have been proposed to tackle impulsive noise, however.

C. Contribution

In this paper, we propose a novel OFDM receiver that performs near-optimally in the presence of impulsive noise while maintaining the $O(N \log N)$ complexity order of the conventional N-tone OFDM receiver. Our approach is based on computing *joint soft* estimates of the channel taps, the impulse-noise samples, the finite-alphabet symbols, and the unknown bits. Moreover, *all* observed tones (i.e., pilots, nulls, and data tones) are exploited in this joint estimation. To do this, we leverage recent work on "generalized approximate message passing" (GAMP) [31], its "turbo" extension to larger factor graphs [34], and off-the-shelf soft-input/soft-output (SISO) decoding [35]. The receiver we propose can be categorized as an extension of the factor-graph-based receiver [30] that explicitly addresses the presence of impulsive noise. The resulting receiver provides a flexible performance-versus-complexity tradeoff and can be parallelized, making it suitable for FPGA implementations [36].

D. Organization and Notation

In Section II, we describe our OFDM, channel, and noise models, and provide an illustrative example of impulsive noise. Then, in Section III, we detail our proposed approach, which we henceforth refer to as joint channel, impulse, symbol, and bit (JCISB) estimation. In Section IV, we provide extensive numerical results, and, in Section V, we conclude the paper.

Notation: Vectors and matrices are denoted by boldface lower-case ($\mathbf{x}$) and upper-case notation ($\mathbf{X}$), respectively. $\mathbf{X}_{\mathcal{R},\mathcal{C}}$ then represents the sub-matrix constructed from rows $\mathcal{R}$ and columns $\mathcal{C}$ of $\mathbf{X}$, where the simplified notation $\mathbf{X}_{\mathcal{R}}$ means $\mathbf{X}_{\mathcal{R},:}$ and ":" indicates all columns of $\mathbf{X}$. The notations $(\cdot)^T$ and $(\cdot)^*$ denote transpose and conjugate transpose, respectively. The probability density function (pdf) of a random variable (RV) X is denoted by $p_X(x)$, with the subscript omitted when clear from the context. Similarly, for discrete RVs, the probability mass function (pmf) is denoted by $P_X(x)$. We denote the circular Gaussian distribution with mean μ and variance γ by $\mathfrak{N}(\mu, \gamma)$ and the pdf of a RV X corresponding to that distribution by $\mathfrak{N}(x; \mu, \gamma)$. The expectation and variance of a RV are then given by $\mathbb{E}\{\cdot\}$ and $\mathbb{V}\{\cdot\}$, respectively. We use the *sans-serif* font to indicate frequency domain variables like X and bold *sans-serif* to indicate frequency domain vectors like $\mathbf{X}$.

II. System Model

A. Coded OFDM Model

We consider an OFDM system with N tones partitioned into N_{p} pilot tones (indexed by the set $\mathcal{P}$), N_{n} null tones (indexed by the set $\mathcal{N}$), and N_{d} data tones (indexed by the set $\mathcal{D}$), each

[1]The approach was shown to be near-optimal in the sense of achieving [32] the pre-log factor of the sparse channel's noncoherent capacity [33].

modulated by a finite-alphabet symbol chosen from an 2^M-ary constellation $\mathbb{S}$. The coded bits (which determine the data symbols) are generated by encoding M_{i} information bits using a rate-R coder, interleaving them, and allocating the resulting $M_{\mathsf{c}} = M_{\mathsf{i}}/R$ bits among an integer number $Q = \lceil M_{\mathsf{c}}/N_{\mathsf{d}}M \rceil$ of OFDM symbols.

In the sequel, we use $\mathsf{S}^{(i)} \in \mathbb{S}$ with $i \in \{1, \ldots, 2^M\}$ to denote the ith element of $\mathbb{S}$, and $\mathbf{c}^{(i)} = [c_1^{(i)}, \ldots, c_M^{(i)}]^T$ to denote the corresponding bits as defined by the symbol mapping. Likewise, we use $\mathsf{S}_k[q]$ to denote the scalar symbol transmitted on the kth tone of the qth OFDM symbol. Based on the tone partition, we note that: $\mathsf{S}_k[q] = \mathsf{p}$ for all $k \in \mathcal{P}$, where $\mathsf{p} \in \mathbb{C}$ is a known pilot symbol; $\mathsf{S}_k[q] = 0$ for all $k \in \mathcal{N}$; and $\mathsf{S}_k[q] = \mathsf{S}^{(l)}$ for some l such that $\mathbf{c}_k[q] = \mathbf{c}^{(l)}$ for all $k \in \mathcal{D}$, where $\mathbf{c}_k[q] = [c_{k,1}[q], \ldots, c_{k,M}[q]]^T$ denotes the coded/interleaved bits corresponding to $\mathsf{S}_k[q]$. On the frame level, we use $\mathbf{c}[q]$ to denote the coded/interleaved bits allocated to the data tones of the qth OFDM symbol, and $\mathbf{c} = [\mathbf{c}[1], \ldots, \mathbf{c}[Q]]$ to denote the entire codeword obtained from the information bits $\mathbf{b} = [b_1, \ldots, b_{M_{\mathsf{i}}}]^T$ by coding/interleaving. Similarly, we use $\mathbf{S}[q] = [\mathsf{S}_0[q], \ldots, \mathsf{S}_{N-1}[q]]^T$ to denote the qth OFDM symbol's tone vector, including pilot, null, and data tones.

For OFDM modulation, an inverse of the unitary N-point discrete Fourier transform (IDFT) matrix $\mathbf{F}$ is applied to the qth OFDM symbol's tone vector $\mathbf{S}[q]$, producing the time-domain sequence $\mathbf{F}^*\mathbf{S}[q] = \boldsymbol{s}[q] = [\boldsymbol{s}_0[q], \ldots, \boldsymbol{s}_{N-1}[q]]^T$, to which a cyclic prefix is prepended. The resulting sequence propagates through an L-tap linear-time-invariant channel with impulse response $\mathbf{h}[q] = [h_0[q], \ldots, h_{L-1}[q]]^T$ before being corrupted by both AWGN and impulsive noise. Assuming a cyclic prefix of length $L-1$, inter-symbol interference is avoided by simply discarding the cyclic prefix at the receiver, after which the remaining N samples are

$$\mathbf{y}[q] = \mathbf{H}[q]\mathbf{s}[q] + \mathbf{n}[q] = \mathbf{H}[q]\mathbf{F}^*\mathbf{S}[q] + \mathbf{n}[q] \tag{1}$$

where $\mathbf{n}[q]$ is the time-domain noise vector and $\mathbf{H}[q]$ is the circulant matrix formed by $\mathbf{h}[q]$ [1]. Applying a DFT, the resulting frequency-domain received vector becomes

$$\mathbf{Y}[q] = \mathbf{F}\mathbf{H}[q]\mathbf{F}^*\mathbf{S}[q] + \mathbf{F}\mathbf{n}[q] = \mathbf{H}[q] \circ \mathbf{S}[q] + \mathbf{N}[q] \tag{2}$$

where $\mathbf{H}[q] = \sqrt{N}\mathbf{F}_{:,1:L}\mathbf{h}[q]$ is the frequency-domain channel vector, $\mathbf{N}[q] = \mathbf{F}\mathbf{n}[q]$ is the frequency-domain noise vector, and $\circ$ denotes the Hadamard (i.e., elementwise) product. The second equality in (2) follows from the fact that a circulant matrix is diagonalized by the Fourier basis. In fact, (2) illustrates the principal advantage of OFDM: each transmitted tone $\mathsf{S}_k[q]$ experiences a flat scalar subchannel, since

$$\mathsf{Y}_k[q] = \mathsf{H}_k[q]\mathsf{S}_k[q] + \mathsf{N}_k[q], \quad \forall k \in \{0, \ldots, N-1\}. \tag{3}$$

B. Channel Modeling

We assume that the channel taps remain constant during the entire duration of one OFDM symbol, as required by (2). Since we make no assumptions on how the taps change across symbols, for simplicity we take $\mathbf{h}[q]$ and $\mathbf{h}[q']$ to be statistically independent for $q \neq q'$. Furthermore, we use the Rayleigh-fading uncorrelated-scattering model

$$\{h_j[q]\}_{q=-\infty}^{\infty} \sim \text{i.i.d } \mathfrak{N}(0, \nu_j) \tag{4}$$

where $[\nu_0, \ldots, \nu_{L-1}]^T = \boldsymbol{\nu}$ is the power delay profile. Extensions to sparse [32], structured-sparse [30], and time-varying sparse channels [37] are straightforward, but not covered here.

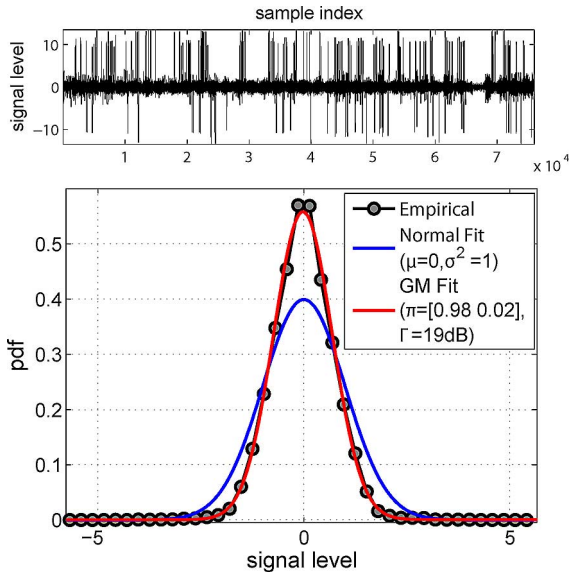


Fig. 1. Modeling a noise realization collected from a receiver embedded in a laptop: A 2-component GM model provides a significantly better fit than a Gaussian model.

C. Impulsive Noise Models

In many wireless and power-line communication (PLC) systems, the additive noise is non-Gaussian (see the example in Fig. 1) and the result of random emission events from uncoordinated interferers (due to, e.g., aggressive spectrum reuse) or non-communicating electronic devices. In his pioneering work, Middleton modeled these random spatio-temporal emissions, or the "noise field," using Poisson point processes (PPP), giving rise to the "Middleton class-A" and "Middleton class-B" noise models (for a recent review see [38]). Recently, his approach has been extended to modeling fields of interferers in wireless and PLC networks using spatial and temporal PPPs, and the resulting interference was shown to follow either the Symmetric alpha stable, the Middleton class-A (MCA), or the more general Gaussian mixture (GM) distribution, depending on the network architecture [9], [10], [39], [40]. Fig. 1 illustrates that a GM model provides a significantly better fit to a noise realization collected from a receiver embedded in a laptop than a Gaussian model does.

Since our factor-graph-based receiver is inherently Bayesian, these statistical models provide natural priors on the noise. Thus, we model the additive noise using a GM model, noting

that—given the pdf parameters—there is no distinction between the MCA and GM models. In particular, we decompose[2] a given time-domain noise sample $n_t = g_t + i_t$ into a Gaussian background component $g_t \sim \mathfrak{N}(0, \gamma^{(0)})$ and a sparse impulsive component i_t with Bernoulli-GM pdf

$$p(i_t) = \pi^{(0)}\delta(i_t) + \sum_{k=1}^{K-1} \pi^{(k)} \mathfrak{N}\left(i_t; 0, \gamma^{(k)}\right) \tag{5}$$

where $\delta(\cdot)$ denotes the Dirac delta and $\sum_{k=0}^{K-1} \pi^{(k)} = 1$. Equivalently, we can model the (hidden) mixture state $z_t \in \{0, \ldots, K-1\}$ of the impulsive component i_t as a random variable, giving rise to the hierarchical model (with $\gamma^{(0)} = 0$)

$$p(i_t|z_t = k) = \mathfrak{N}\left(i_t; 0, \gamma^{(k)}\right) \tag{6a}$$

$$P(z_t = k) = \pi^{(k)}. \tag{6b}$$

In many applications, such as PLC, the noise is not only impulsive but also *bursty* and thus the noise samples are no longer statistically independent. Such burstiness can be captured via a Bernoulli-Gaussian[3] hidden Markov model (BGHMM) on the impulse-noise $\{i_t\}$ [8], [41] or equivalently a Markov model on the GM state z_t in (6). For this, we model the sequence $\{z_t\}$ as a homogeneous (stationary) K-ary Markov chain with a state transition matrix $\mathbf{T}$ such that

$$\mathbf{T}_{i,j} = P(z_t = j|z_{t-1} = i) \quad \forall i, j \in \{0, \ldots, K-1\}. \tag{7}$$

In this case, the marginal pmf $\boldsymbol{\pi} = [\pi_0, \ldots, \pi_{K-1}]$ of steady-state z_t obeys $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{T}$, and the mean duration of the event $z = k$ is $1/(1 - \mathbf{T}_{k,k})$ [41].

As an illustrative example, Fig. 2 plots two realizations of the total noise $\{n_t\}$ with impulsive component $\{i_t\}$ generated by the hierarchical Bernoulli-GM model (6). Both realizations have identical marginal statistics: their impulsive components have two non-trivial emission states with powers 20 dB and 30 dB above the background noise power that occur 7% and 3% of the time, respectively. However, in one case the emission state $\{z_t\}$ was generated i.i.d whereas in the other case it is generated Markov with state-transition matrix

$$\mathbf{T} = \begin{bmatrix} 0.989 & 0.006 & 0.005 \\ 0.064 & 0.857 & 0.079 \\ 0.183 & 0.150 & 0.667 \end{bmatrix}. \tag{8}$$

The GHMM realization clearly exhibits bursty behavior.

In practice, assuming the noise statistics are slowly varying, the noise parameters $\{\gamma_k\}_{k=0}^{K-1}$ and $\mathbf{T}$ can be estimated using the expectation-maximization (EM) algorithm [26] during quiet intervals when there is no signal transmission.

[2]Our approach is equivalent to modeling the total noise n_t by a GM pdf $(n_t) = \sum_{k=0}^{K-1} \pi^{(k)} \mathfrak{N}(n_t; 0, \nu^{(k)})$ with $\nu^{(0)} = \gamma^{(0)}$ and $\nu^{(k)} = \gamma^{(0)} + \gamma^{(k)}$ for $k > 1$.

[3]Here, Bernoulli refers to the impulse noise support. The non-zero samples follow multiple Gaussian emission probabilities characterized by different variances $\gamma^{(k)}$.

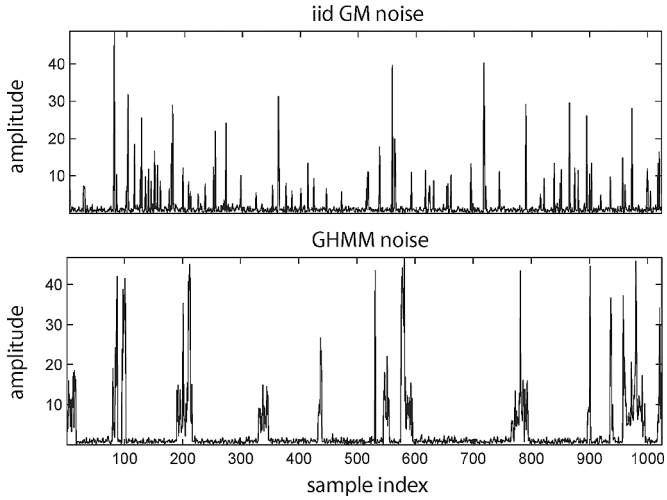


Fig. 2. Two realizations of noise $\{n_t\}$ with identical marginals but different temporal statistics: the emission states $\{z_t\}$ in the top were generated i.i.d and those in the bottom were generated Markov via (8) to model burstiness.

III. Message-Passing Receiver Design

In this section, we design computationally efficient message-passing receivers that perform near-optimal bit decoding which, as we shall see, involves jointly estimating the coded bits, finite-alphabet symbols, channel taps, and impulsive noise samples. In doing so, our receivers exploit knowledge of the statistical channel and noise models discussed above and the OFDM signal structure (i.e., the pilot and null tones, the finite-alphabet symbol constellation, and the codebook).

A. MAP Decoding of OFDM in Impulsive Noise

Maximum a posteriori (MAP) decoding, i.e.,

$$\hat{b}_m = \underset{b_m \in \{0,1\}}{\arg\max}\, P(b_m|\mathbf{Y}) \quad \forall m \in \{1, \ldots, M_{\mathrm{i}}\} \tag{9}$$

is well known to be optimal in the sense of minimizing the bit-error rate. Here, $\mathbf{Y} = [\mathbf{Y}[1], \ldots, \mathbf{Y}[Q]]$ collects the received OFDM symbols of the corresponding frame. Using the law of total probability, we can write the posterior information-bit probability from (9) as

$$P(b_m|\mathbf{Y}) = \sum_{\mathbf{b}_{\setminus \mathbf{m}}} P(\mathbf{b}|\mathbf{Y}) \propto \sum_{\mathbf{b}_{\setminus \mathbf{m}}} p(\mathbf{Y}|\mathbf{b}) P(\mathbf{b}) \tag{10}$$

$$\propto \sum_{\mathsf{S},\mathbf{c},\mathbf{b}_{\setminus \mathbf{m}}} \prod_{q=1}^{Q} \int_{\mathbf{i}[q],\mathbf{h}[q]} p(\mathbf{Y}[q]|\mathbf{h}[q], \mathbf{i}[q], \mathbf{S}[q]) \times p(\mathbf{h}[q]) p(\mathbf{i}[q]) P(\mathbf{S}|\mathbf{c}) P(\mathbf{c}|\mathbf{b}) \tag{11}$$

$$= \sum_{\mathsf{S},\mathbf{c},\mathbf{z},\mathbf{b}_{\setminus \mathbf{m}}} \prod_{q} \int_{\mathbf{i}[q],\mathbf{h}[q]} \prod_{k=0}^{N-1} p(\mathsf{Y}_k[q]|\mathsf{S}_k[q], \mathbf{h}[q], \mathbf{i}[q]) \times P(\mathsf{S}_k[q]|\mathbf{c}_k[q]) \times \prod_{j=L}^{L+N-1} p(i_j[q]|z_j[q]) P(z_j[q]|z_{j-1}[q]) \times \prod_{l=0}^{L-1} p(h_l[q]) P(\mathbf{c}|\mathbf{b}) \tag{12}$$

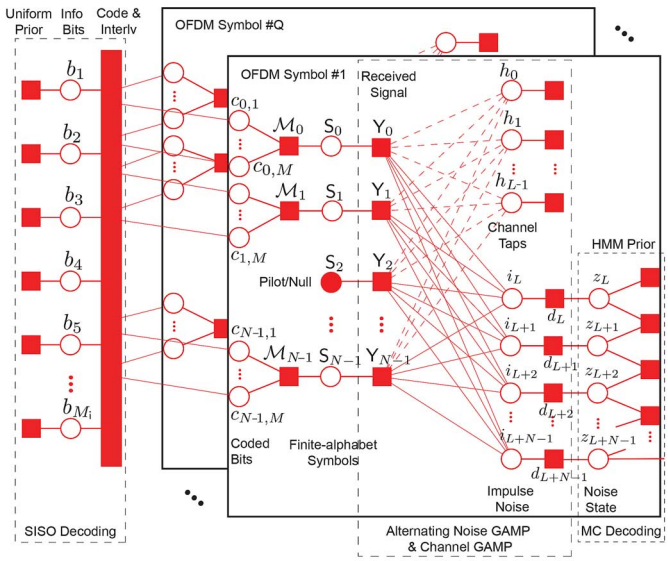


Fig. 3. Factor graph representation of a coded data frame spanning Q OFDM symbols. Round open circles denote random variables, round solid circles denote deterministic variables (e.g., known pilots or nulls), and solid squares denote pdf factors. The large rectangle on the left represents the coding-and-interleaving subgraph, whose details are immaterial. The time-domain impulse-noise quantities i_t and z_t start at time $t = L$ due to the use of an $L-1$-length cyclic prefix.

where "$\propto$" denotes equality up to a constant, $\mathbf{b}_{\backslash \mathbf{m}} = [b_1, \ldots, b_{m-1}, b_{m+1}, \ldots, b_{M_i}]^T$, and the information bits are assumed to be independent with $P(b_m) = 1/2 \ \forall m$. Equation (12) shows that optimal decoding of b_m involves marginalizing over the finite-alphabet symbols $\mathbf{S}$, coded bits $\mathbf{c}$, noise states $\mathbf{z}$, impulse noise samples $\mathbf{i}$, channel taps $\mathbf{h}$, and other info bits $\mathbf{b}_{\backslash \mathbf{m}}$.

The probabilistic structure exposed by the factorization (12) is illustrated by the *factor graph* in Fig. 3. There and in the sequel, for brevity, we drop the "q" (i.e., OFDM symbol) index when doing so does not cause confusion. Note that we write the factor graph using time-domain channel coefficients, rather than frequency-domain channel coefficients, because we assume that the time-domain coefficients are statistically independent, as is standard practice in the literature. Noting that the corresponding frequency-domain coefficients are *not* statistically independent, there is no advantage to writing the factor graph using frequency-domain channel coefficients. Later in Section III.C, we show that, despite the dense subgraph connecting the time-domain channel taps, efficient low-complexity channel inference is possible.

Clearly, direct evaluation of $P(b_m|\mathbf{Y})$ from (12) is computationally intractable due to the high-dimensional integrals involved. Belief propagation (BP), and in particular the sum-product algorithm (SPA) [26] described below, offers a practical alternative to direct computation of marginal posteriors. In fact, when the factor graph has no loops, the SPA performs exact inference after only two rounds of message passing (i.e., forward and backward). On the other hand, when the factor graph is loopy, the computation of exact marginal posteriors is generally NP hard [42] and thus the posteriors computed by BP are generally inexact. Nevertheless, loopy BP has been successfully applied to many important problems, such as multi-user detection [43], [44], turbo decoding [45], LDPC decoding [35], compressed sensing [31], [46], and others.

In fact, for certain large dense loopy graphs that arise in the context of compressed sensing, SPA approximations such as the AMP [46] and GAMP [31] algorithms are known to obey a state evolution whose fixed points, when unique, yield exact posteriors [47], [48]. Looking at the factor graph in Fig. 3, we see dense loopy sub-graphs between the factors $\{\mathsf{Y}_k\}$ and the time-domain noise samples $\{i_t\}$ and channel taps $\{h_l\}$, which are due to the linear mixing of the Fourier matrix $\mathbf{F}$. It is these types of dense loopy graphs for which AMP and GAMP are designed.[4] In the sequel, we will describe exactly how we combine the sum-product and GAMP algorithms for approximate computation of the bit posteriors in (12). First, however, we review the SPA.

B. Belief Propagation Using Sum-Product Algorithm

Belief propagation (BP) transforms a high-dimensional marginalization problem (like (12)) into a series of local low-dimensional marginalization problems by passing beliefs, or messages, which usually take the form of (possibly un-normalized) pdfs or pmfs, along the edges of a factor graph. The sum-product algorithm (SPA) [26] is probably the best known approach to BP, and it operates according to the following rules (note that integrals are replaced by sums for discrete random variables):

1) Messages From Factor Nodes to Variables: Suppose the pdf factor $f_s(x_1, \ldots, x_L)$ depends on the variables $\mathcal{X}_s = \{x_l\}_{l=1}^L$. Then the message passed from factor node f_s to variable node $x_m \in \mathcal{X}_s$ is

$$\mu_{f_s \to x_m}(x_m) = \int_{\{x_l\}_{l \neq m}} f_s(x_1, \ldots, x_L) \prod_{l \neq m} \mu_{x_l \to f_s}(x_l)$$

representing the belief that node f_s has about the variable x_m.

2) Messages From Variables to Factor Nodes: Suppose the factors $\mathcal{F}_m = \{f_1, \ldots, f_M\}$ all involve the variable x_m. Then the message passed from variable node x_m to factor node $f_s \in \mathcal{F}_m$ is

$$\mu_{x_m \to f_s}(x_m) = \prod_{k=1}^{M} \mu_{f_k \to x_m}(x_m)$$

and represents the belief about the variable x_m that node x_m passes to node f_s.

3) Marginal Beliefs: SPA's approximation to the marginal posterior pdf on the variable x_m is

$$p(x_m) = C \prod_{f_s \in \mathrm{ne}(x_m)} \mu_{f_s \to x_m}(x_m)$$

where C is a normalization constant and $\mathrm{ne}(x_m)$ is the set of all x_m neighbors.

[4]Although rigorous GAMP guarantees have been established only for generalized linear inference problems with i.i.d sub-Gaussian transform matrices [48], equally good performance has been empirically observed over much wider classes of matrices [49].

C. Joint Channel/Impulse-Noise Estimation and Decoding

We now propose a strategy to approximate the bit posteriors in (12) by iterating (an approximation of) the SPA on the loopy factor graph in Fig. 3. To distinguish our approach from others in the literature, we will refer to it as "joint channel, impulse, symbol, and bit estimation" (JCISB).

Given the loopy nature of the factor graph, there exists considerable freedom in the message-passing schedule. In JCISB, we choose to repeatedly pass messages from right to left, and then left to right, as follows.

1) Beliefs about coded bits $\{c_{k,m}\}$ flow rightward through the symbol-mapping nodes $\{\mathcal{M}_k\}$, the finite-alphabet symbol nodes $\{\mathsf{S}_k\}$, and into the factor nodes $\{\mathsf{Y}_k\}$.
2) GAMP-based messages are then passed repeatedly between the $\{\mathsf{Y}_k\}$ and $\{h_l\}$ nodes until convergence.
3) GAMP-based messages are passed repeatedly between the $\{\mathsf{Y}_k\}$ and $\{i_t\}$ nodes until convergence, and then through the $\{z_t\}$ nodes using the forward-backward algorithm, alternating these two steps until convergence.
4) Finally, the messages are propagated from $\{\mathsf{Y}_k\}$ leftward through the symbol nodes $\{\mathsf{S}_k\}$, the symbol-mapping nodes $\{\mathcal{M}_k\}$, the coded-bit nodes $\{c_{k,m}\}$, and the coding-interleaving block—the last step via an off-the-shelf soft-input/soft-output (SISO) decoder.

In the sequel, we will refer to steps 1)–4) as a "turbo" iteration, and to the iterations within step 3) as "impulse iterations," We note that it is also possible to execute a parallel schedule if the hardware platform supports it. The details of these four message passing steps are given below.

1) Bits to Symbols: Beliefs about the coded bits $\{c_{k,m}\}_{m=1}^{M}$ (for each data tone $k \in \mathcal{D}$) are first passed through the symbol mapping factor node $\mathcal{M}_k$. The SPA dictates that

$$\begin{aligned}\mu_{\mathcal{M}_k\to\mathsf{S}_k}\left(\mathsf{S}^{(i)}\right) &= \sum_{\mathbf{c}_k\in\{0,1\}^M} P\left(\mathsf{S}^{(i)}|\mathbf{c}_k\right)\prod_{m=1}^{M}\mu_{c_{k,m}\to\mathcal{M}_k}(c_{k,m})\\ &= \prod_{m=1}^{M}\mu_{c_{k,m}\to\mathcal{M}_k}\left(c_m^{(i)}\right) \qquad (13)\end{aligned}$$

where (13) follows from the deterministic symbol mapping $P(\mathsf{S}^{(i)}|\mathbf{c}^{(j)}) = \delta_{i-j}$. The resulting message is then copied forward through the S_k node, i.e., $\mu_{\mathsf{S}_k\to\mathsf{Y}_k} = \mu_{\mathcal{M}_k\to\mathsf{S}_k}$, also according to the SPA. Note that, at the start of the first turbo iteration, we have no knowledge of the bits and thus we take $\mu_{c_{k,m}\to\mathcal{M}_k}(c)$ to be uniform across $c\in\{0,1\}$ for all m,k.

2) GAMP for Channel Estimation: The next step in our message-passing schedule is to pass messages between the factor nodes $\{\mathsf{Y}_k\}$ and the time-domain channel nodes $\{h_l\}$. According to the SPA, the message passed from Y_k to h_l is

$$\begin{aligned}\mu_{\mathsf{Y}_k\to h_l}(h_l) = \sum_{\mathsf{S}_k}\int_{\mathbf{i},\mathbf{h}_{\setminus l}} & p(\mathsf{Y}_k|\mathsf{S}_k,\mathbf{h},\mathbf{i})\mu_{\mathsf{S}_k\to\mathsf{Y}_k}(\mathsf{S}_k)\\ &\times\prod_{z\neq l}\mu_{h_z\to\mathsf{Y}_k}(h_z)\prod_j\mu_{i_j\to\mathsf{Y}_k}(i_j). \qquad (14)\end{aligned}$$

Exact evaluation of (14) involves an integration of the same high-dimensionality (N) that made (12) intractable. Thus, we instead approximate the message passing between the $\{\mathsf{Y}_k\}$ and $\{h_l\}$ nodes using *generalized approximate message passing* (GAMP) algorithm [31] reviewed in Appendix A and summarized in Table I.

TABLE I
THE GAMP$(\mathbf{y},\mathbf{z},\mathbf{\Phi},\mathbf{x})$ ALGORITHM WITH COEFFICIENTS-OF-INTEREST $\mathbf{x}$, LINEAR TRANSFORM $\mathbf{\Phi}$, TRANSFORM OUTPUTS $\mathbf{z}=\mathbf{\Phi x}$, AND OBSERVED OUTPUTS $\mathbf{y}$

inputs: $\{p(x_j)\}_{\forall j},\{p(y_i|z_i)\}_{\forall i},\mathbf{y},\mathbf{\Phi},T_{\max}$

definitions:

$$p(z_i|\mathbf{y};\hat{p},\gamma^p) = \frac{p(y_i|z_i)\,\mathcal{N}(z_i;\hat{p},\gamma^p)}{\int_{z_i'}p(y_i|z_i')\,\mathcal{N}(z_i';\hat{p},\gamma^p)} \qquad (D1)$$

$$p(x_j|\mathbf{y};\hat{r},\gamma^r) = \frac{p(x_j)\,\mathcal{N}(x;\hat{r},\gamma^r)}{\int_{x_j'}p(x_j')\,\mathcal{N}(x_j';\hat{r},\gamma^r)} \qquad (D2)$$

initialize:

$$\forall j:\hat{x}_j(1) = \textstyle\int_{x_j} x_j\,p(x_j) \qquad (I1)$$

$$\forall j:\gamma_j^x(1) = \textstyle\int_{x_j}|x_j-\hat{x}_j(1)|^2p(x_j) \qquad (I2)$$

$$\forall i:\hat{s}_i(0) = 0 \qquad (I3)$$

for $t=1:T_{\max}$,

$$\forall i:\gamma_i^p(t) = \textstyle\sum_{j=1}^{N}|\Phi_{ij}|^2\gamma_j^x(t) \qquad (R1)$$

$$\forall i:\hat{p}_i(t) = \textstyle\sum_{j=1}^{N}\Phi_{ij}\hat{x}_j(t)-\gamma_i^p(t)\,\hat{s}_i(t-1) \qquad (R2)$$

$$\forall i:\gamma_i^z(t) = \mathbb{V}\{z_i|\mathbf{y};\hat{p}_i(t),\gamma_i^p(t)\} \qquad (R3)$$

$$\forall i:\hat{z}_i(t) = \mathbb{E}\{z_i|\mathbf{y};\hat{p}_i(t),\gamma_i^p(t)\} \qquad (R4)$$

$$\forall i:\gamma_i^s(t) = \big(1-\gamma_i^z(t)/\gamma_i^p(t)\big)/\gamma_i^p(t) \qquad (R5)$$

$$\forall i:\hat{s}_i(t) = \big(\hat{z}_i(t)-\hat{p}_i(t)\big)/\gamma_i^p(t) \qquad (R6)$$

$$\forall j:\gamma_j^r(t) = \big(\textstyle\sum_{i=1}^{M}|\Phi_{ij}|^2\gamma_i^s(t)\big)^{-1} \qquad (R7)$$

$$\forall j:\hat{r}_j(t) = \hat{x}_j(t)+\gamma_j^r(t)\textstyle\sum_{i=1}^{M}\Phi_{ij}^*\hat{s}_i(t) \qquad (R8)$$

$$\forall j:\gamma_j^x(t+1) = \mathbb{V}\{x_j|\mathbf{y};\hat{r}_j(t),\gamma_j^r(t)\} \qquad (R9)$$

$$\forall j:\hat{x}_j(t+1) = \mathbb{E}\{x_j|\mathbf{y};\hat{r}_j(t),\gamma_j^r(t)\} \qquad (R10)$$

outputs: $\{\hat{z}_i(T),\gamma_i^z(T)\}_{\forall i},\{\hat{r}_j(T),\gamma_j^r(T),\hat{x}_j(T),\gamma_j^x(T)\}_{\forall j}$

To do this, we temporarily treat the messages $\{\mu_{\mathsf{S}_k\to\mathsf{Y}_k}\}$ and $\{\mu_{i_j\to\mathsf{Y}_k}\}$ as fixed, allowing us to employ "GAMP$(\mathsf{Y},\mathsf{H},\sqrt{N}\mathbf{F},\mathbf{h})$," using the notation established in the caption of Table I. Definition (D1) [later used in steps (R3)–(R4)] requires us to specify the likelihood $p(\mathsf{Y}_k|\mathsf{H}_k)$ relating the transform output H_k to the corresponding observed output Y_k. From Fig. 3, we see that there are two types of belief flowing into each Y_k node (apart from beliefs about $\{h_l\}$) that determine this likelihood: beliefs about the symbols $\{\mathsf{S}_k\}$, which we parameterize as $\boldsymbol{\beta}_k = [\beta_k^{(1)},\ldots,\beta_k^{(|\mathbb{S}|)}]$ with $\beta_k^{(i)} = \mu_{\mathsf{S}_k\to\mathsf{Y}_k}(\mathsf{S}^{(i)})$, and beliefs about the frequency-domain impulsive noise $\{\mathsf{I}_k\}$, which GAMP approximates as $\mathfrak{N}(\mathsf{I}_k;\hat{\mathsf{I}}_k,\gamma_k^{\mathsf{I}})$, where the values $\{\hat{\mathsf{I}}_k,\gamma_k^{\mathsf{I}}\}$ were computed by GAMP$(\mathsf{Y},\mathsf{I},\mathbf{F},\mathbf{i})$ in the previous turbo iteration.[5] Here, $\mathsf{I}=\mathbf{Fi}$ refers to the impulsive component of the frequency-domain noise $\mathsf{N}=\mathsf{I}+\mathsf{G}$, with $\{\mathsf{G}_k\}\sim$ i.i.d $\mathfrak{N}(0,\gamma^{(0)})$, so that [from (2)]

$$\mathsf{Y} = \mathsf{H}\circ\mathsf{S}+\mathsf{I}+\mathsf{G}. \qquad (15)$$

From (3) and (15), the GAMP$(\mathsf{Y},\mathsf{H},\sqrt{N}\mathbf{F},\mathbf{h})$ likelihood is

$$p(\mathsf{Y}_k|\mathsf{H}_k) = \begin{cases}\mathfrak{N}\left(\mathsf{Y}_k;\mathsf{p}\mathsf{H}_k+\hat{\mathsf{I}}_k,\gamma_k^{\mathsf{I}}+\gamma^{(0)}\right) & k\in\mathcal{P}\\ \sum_{l=1}^{|\mathbb{S}|}\beta_k^{(l)}\mathfrak{N}\left(\mathsf{Y}_k;\mathsf{S}^{(l)}\mathsf{H}_k+\hat{\mathsf{I}}_k,\gamma_k^{\mathsf{I}}+\gamma^{(0)}\right) & k\in\mathcal{D}\end{cases} \qquad (16)$$

with the corresponding "output MMSE estimation functions" $\mathbb{E}\{\mathsf{H}_k|\mathsf{Y};\hat{p},\gamma^p\}$ and $\mathbb{V}\{\mathsf{H}_k|\mathsf{Y};\hat{p},\gamma^p\}$, as used in

[5] During the first turbo iteration, we use $\hat{\mathsf{I}}_k=0$ and $\gamma_k^{\mathsf{I}}=\gamma^i\forall k$.

TABLE II
GAMP OUTPUT MMSE ESTIMATION FUNCTIONS USED IN JCISB

Tone Type	GAMP$(\mathbf{Y},\mathbf{H},\sqrt{N}\mathbf{F},\mathbf{h})$	
	$\mathbb{E}\{\mathsf{H}_k\|\mathbf{Y};\hat{p},\gamma^p\}$	$\mathbb{V}\{\mathsf{H}_k\|\mathbf{Y};\hat{p},\gamma^p\}$
Pilot: $k\in\mathcal{P}$	$\hat{p}+\gamma^p\mathsf{p}^*(\mathsf{Y}_k-\hat{\mathsf{I}}_k-\mathsf{p}\hat{p})/(\gamma^{(0)}+\gamma_k^{\mathsf{I}}+\rho_{\mathsf{p}}\gamma^p)$	$\gamma^p(\gamma^{(0)}+\gamma_k^{\mathsf{I}})/(\gamma^{(0)}+\gamma_k^{\mathsf{I}}+\rho_{\mathsf{p}}\gamma^p)$
Data: $k\in\mathcal{D}$	$\hat{p}+\sum_{l=1}^{\|\mathbb{S}\|}\lambda_k^{(l)}\frac{\gamma^p\mathsf{S}^{*(l)}(\mathsf{Y}_k-\hat{\mathsf{I}}_k-\hat{p}\mathsf{S}^{(l)})}{(\gamma^{(0)}+\gamma_k^{\mathsf{I}}+\|\mathsf{S}^{(l)}\|^2\gamma^p)}$ where $\lambda_k^{(l)}=p(\mathsf{Y}_k\|\mathsf{S}^{(l)})\beta_k^{(l)}/\sum_j p(\mathsf{Y}_k\|\mathsf{S}^{(j)})\beta_k^{(j)}$ and $p(\mathsf{Y}_k\|\mathsf{S}^{(l)})=\mathfrak{N}(\mathsf{Y}_k;\hat{\mathsf{I}}_k+\hat{p}\mathsf{S}^{(l)},\gamma^{(0)}+\gamma_k^{\mathsf{I}}+\|\mathsf{S}^{(l)}\|^2\gamma^p)$	$\sum_{l=1}^{\|\mathbb{S}\|}\lambda_k^{(l)}\Big[\frac{\gamma^p(\gamma^{(0)}+\gamma_k^{\mathsf{I}})}{\gamma^{(0)}+\gamma_k^{\mathsf{I}}+\|\mathsf{S}^{(l)}\|^2\gamma^p}+\Big\|\hat{p}+\frac{\gamma^p\mathsf{S}^{*(l)}(\mathsf{Y}_k-\hat{\mathsf{I}}_k-\hat{p}\mathsf{S}^{(l)})}{(\gamma^{(0)}+\gamma_k^{\mathsf{I}}+\|\mathsf{S}^{(l)}\|^2\gamma^p)}\Big\|^2\Big]-\|\mathbb{E}\{\mathsf{H}_k\|\mathbf{Y};\hat{p},\gamma^p\}\|^2$

Tone Type	GAMP$(\mathbf{Y},\mathbf{I},\mathbf{F},\mathbf{i})$	
	$\mathbb{E}\{\mathsf{I}_k\|\mathbf{Y};\hat{p},\gamma^p\}$	$\mathbb{V}\{\mathsf{I}_k\|\mathbf{Y};\hat{p},\gamma^p\}$
Null: $k\in\mathcal{N}$	$(\gamma^p\mathsf{Y}_k-\gamma^{(0)}\hat{p})/(\gamma^{(0)}+\gamma^p)$	$\gamma^p\gamma^{(0)}/(\gamma^{(0)}+\gamma^p)$
Pilot: $k\in\mathcal{P}$	$\hat{p}+\gamma^p(\mathsf{Y}_k-\hat{p}-\hat{\mathsf{H}}_k\mathsf{p})/(\gamma^{(0)}+\gamma^p+\rho_{\mathsf{p}}\gamma_k^{\mathsf{H}})$	$\gamma^p(\gamma^{(0)}+\rho_{\mathsf{p}}\gamma_k^{\mathsf{H}})/(\gamma^{(0)}+\gamma^p+\rho_{\mathsf{p}}\gamma_k^{\mathsf{H}})$
Data: $k\in\mathcal{D}$	$\hat{p}+\sum_{l=1}^{\|\mathbb{S}\|}\lambda_k^{(l)}\frac{\gamma^p(\mathsf{Y}_k-\hat{p}-\hat{\mathsf{H}}_k\mathsf{S}^{(l)})}{(\gamma^{(0)}+\gamma^p+\|\mathsf{S}^{(l)}\|^2\gamma_k^{\mathsf{H}})}$ where $\lambda_k^{(l)}=p(\mathsf{Y}_k\|\mathsf{S}^{(l)})\beta_k^{(l)}/\sum_j p(\mathsf{Y}_k\|\mathsf{S}^{(j)})\beta_k^{(j)}$ and $p(\mathsf{Y}_k\|\mathsf{S}^{(l)})=\mathfrak{N}(\mathsf{Y}_k;\hat{p}+\hat{\mathsf{H}}_k\mathsf{S}^{(l)},\gamma^{(0)}+\gamma^p+\|\mathsf{S}^{(l)}\|^2\gamma_k^{\mathsf{H}})$	$\sum_{l=1}^{\|\mathbb{S}\|}\lambda_k^{(l)}\Big[\frac{\gamma^p(\gamma^{(0)}+\|\mathsf{S}^{(l)}\|^2\gamma_k^{\mathsf{H}})}{(\gamma^{(0)}+\gamma^p+\|\mathsf{S}^{(l)}\|^2\gamma_k^{\mathsf{H}})}+\Big\|\hat{p}+\frac{\gamma^p(\mathsf{Y}_k-\hat{p}-\hat{\mathsf{H}}_k\mathsf{S}^{(l)})}{\gamma^{(0)}+\gamma^p+\|\mathsf{S}^{(l)}\|^2\gamma_k^{\mathsf{H}}}\Big\|^2\Big]-\|\mathbb{E}\{\mathsf{I}_k\|\mathbf{Y};\hat{p},\gamma^p\}\|^2$

steps (R3)–(R4), specified in Table II. (See Appendix B for derivations).

GAMP$(\mathbf{Y},\mathbf{H},\sqrt{N}\mathbf{F},\mathbf{h})$ also requires us to derive the "input MMSE estimation functions" $\mathbb{E}\{h_j|\mathbf{Y},\hat{r},\gamma^r\}$ and $\mathbb{V}\{h_j|\mathbf{Y},\hat{r},\gamma^r\}$ for GAMP steps (R9)–(R10). Given the channel model specified in Section II.B and definition (D2), it is straightforward to show [31] that the input MMSE estimation functions are $\mathbb{E}\{h_j|\mathbf{Y},\hat{r},\gamma^r\}=\nu_j\hat{r}/(\nu_j+\gamma^r)$ and $\mathbb{V}\{h_j|\mathbf{Y},\hat{r},\gamma^r\}=\nu_j\gamma^r/(\nu_j+\gamma^r)$.

After GAMP$(\mathbf{Y},\mathbf{H},\sqrt{N}\mathbf{F},\mathbf{h})$ is iterated to convergence, the outputs $\{\hat{\mathsf{H}}_k\}$ and $\{\gamma_k^{\mathsf{H}}\}$ of steps (R4)–(R3) are close approximations to the marginal posterior mean and variance, respectively, of $\{\mathsf{H}_k\}$. These outputs will be used in the next step of the message-passing schedule, as described below. Similarly, the outputs $\{\hat{h}_l\}$ and $\{\gamma_l^h\}$ of steps (R10)–(R9) are close approximations to the marginal posterior mean and variance, respectively, of $\{h_l\}$.

3) Turbo-GAMP for Noise Estimation: The next step in our schedule is to pass messages between the factor nodes $\{\mathsf{Y}_k\}$, the time-domain impulse-noise nodes $\{i_t\}$, and the noise-state nodes $\{z_t\}$. According to the SPA, the message passed from Y_k to i_t is

$$\mu_{\mathsf{Y}_k\to i_t}(i_t)=\sum_{\mathsf{S}_k}\int_{\mathbf{i}_{\setminus t},\mathbf{h}}p(\mathsf{Y}_k|\mathsf{S}_k,\mathbf{h},\mathbf{i})\mu_{\mathsf{S}_k\to\mathsf{Y}_k}(\mathsf{S}_k)\times\prod_l\mu_{h_l\to\mathsf{Y}_k}(h_l)\prod_{j\neq t}\mu_{i_j\to\mathsf{Y}_k}(i_j)\quad(17)$$

which poses the same difficulties as (12) and (14).

Although GAMP can help approximate the messages in (17), GAMP alone is insufficient due to connections between the $\{d_t\}$ nodes, which are used to model the burstiness of the time-domain impulse-noise $\{i_t\}$. However, recognizing that the underlying problem is estimation of a clustered-sparse sequence $\{i_t\}$ from compressed linear measurements, we can use the solution proposed in [34], which alternated GAMP with the forward-backward (FB) algorithm [26], as described below.

First, by temporarily treating the messages $\{\mu_{d_t\to i_t}\}$, $\{\mu_{\mathsf{S}_k\to\mathsf{Y}_k}\}$, and $\{\mu_{h_l\to\mathsf{Y}_k}\}$ as fixed, we can apply GAMP$(\mathbf{Y},\mathbf{I},\mathbf{F},\mathbf{i})$ under the likelihood model

$$p(\mathsf{Y}_k|\mathsf{I}_k)=\begin{cases}\mathfrak{N}\left(\mathsf{Y}_k;\mathsf{I}_k,\gamma^{(0)}\right) & \text{if } k\in\mathcal{N}\\ \mathfrak{N}\left(\mathsf{Y}_k;\mathsf{p}\hat{\mathsf{H}}_k+\mathsf{I}_k,\rho_{\mathsf{p}}\gamma_k^{\mathsf{H}}+\gamma^{(0)}\right) & \text{if } k\in\mathcal{P}\\ \sum_{l=1}^{\|\mathbb{S}\|}\beta_k^{(l)}\mathfrak{N}\left(\mathsf{Y}_k;\mathsf{S}^{(l)}\hat{\mathsf{H}}_k+\mathsf{I}_k,\rho_{\mathsf{p}}\gamma_k^{\mathsf{H}}+\gamma^{(0)}\right) & \text{if } k\in\mathcal{D}\end{cases}\quad(18)$$

implied by (3) and (15), and the coefficient prior

$$p(i_t)=\pi_t^{(0)}\delta(i_t)+\sum_{k=1}^{K-1}\pi_t^{(k)}\mathfrak{N}\left(i_t;0,\gamma^{(k)}\right)\quad(19)$$

implied by (5). In (18), $\beta_k^{(i)}=\mu_{\mathsf{S}_k\to\mathsf{Y}_k}(\mathsf{S}^{(i)})$ are the symbol beliefs coming from the $\{\mathsf{S}_k\}$ nodes and $\{\hat{\mathsf{H}}_k,\gamma_k^{\mathsf{H}}\}$ are the frequency-domain channel estimates previously calculated by GAMP$(\mathbf{Y},\mathbf{H},\sqrt{N}\mathbf{F},\mathbf{h})$. Meanwhile, in (19), $\{\pi_t^{(k)}\}_{k=0}^{K-1}$ represents the pmf on the noise state z_t that is set as $\pi_t^{(k)}=\mu_{z_t\to d_t}(k)/(\sum_{l=0}^{K-1}\mu_{z_t\to d_t}(l))$. The resulting output MMSE estimation functions, derived in Appendix C, are listed in Table II, and the input MMSE estimation functions are

$$\mathbb{E}\{i_t|\mathbf{Y},\hat{r},\gamma^r\}=\sum_{k=0}^{K-1}\alpha_t^{(k)}\frac{\gamma^{(k)}\hat{r}}{\gamma^{(k)}+\gamma^r}\quad(20)$$

$$\mathbb{V}\{i_t|\mathbf{Y},\hat{r},\gamma^r\}=\sum_{k=0}^{K-1}\frac{\alpha_t^{(k)}}{\gamma^r+\gamma^{(k)}}\left(\gamma^r\gamma^{(k)}+\frac{|\gamma^{(k)}\hat{r}|^2}{\gamma^r+\gamma^{(k)}}\right)-|\mathbb{E}\{i_t|\mathbf{Y},\hat{r},\gamma^r\}|^2.\quad(21)$$

Here, $\{\alpha_t^{(k)}\}_{k=0}^{K-1}$ is the posterior pmf for noise-state z_t, with

$$\alpha_t^{(k)}=P(z_t=k|\hat{r})=\frac{p(\hat{r}|z_t=k)\pi_t^{(k)}}{\sum_{l=0}^{K-1}p(\hat{r}|z_t=l)\pi_t^{(l)}}\quad(22)$$

where $p(\hat{r}|z_t=k)=\mathfrak{N}(\hat{r};0,\gamma^r+\gamma^{(k)})$ is the noise state likelihood.

Using these input and output MMSE estimation functions, $\mathrm{GAMP}(\mathbf{Y},\mathsf{I},\mathbf{F},\mathbf{i})$ is iterated until convergence, generating (for each t) an outgoing belief $\mu_{i_t\to d_t}(i_t)=\mathfrak{N}(i_t;\hat{r},\gamma^r)$ about the noise-impulse i_t. This belief flows through the factor node d_t which, according to the SPA, gives the rightward flowing noise-state belief

$$\mu_{d_t\to z_t}(z_t=k)\propto\mathfrak{N}\left(\hat{r};0,\gamma^r+\gamma^{(k)}\right) \tag{23}$$

that acts as a prior for "Markov-chain (MC) decoding," i.e., inference on the rightmost sub-graph in Fig. 3. Since the MC sub-graph is non-loopy, it suffices to apply one pass of the forward-backward algorithm; see [26] for details. Subsequently the refined noise-state beliefs $\{\mu_{z_t\to d_t}\}$ are passed back to the noise subgraph where each is used to compute the corresponding pmf $\{\pi_t^{(k)}\}_{k=0}^{K-1}$ used in (22) by the next invocation of $\mathrm{GAMP}(\mathbf{Y},\mathsf{I},\mathbf{F},\mathbf{i})$.

When the noise-state beliefs $\{\mu_{z_t\to d_t}\}$ have converged, the impulse-noise iterations are terminated and the $\{\hat{\mathsf{I}}_k,\gamma_k^{\mathsf{I}}\}$ produced by $\mathrm{GAMP}(\mathbf{Y},\mathsf{I},\mathbf{F},\mathbf{i})$ are close approximations to the marginal posterior means and variances of $\{\mathsf{I}_k\}$ that will be used by $\mathrm{GAMP}(\mathbf{Y},\mathbf{H},\sqrt{N}\mathbf{F},\mathbf{h})$ in the next turbo iteration. In addition, for each data tone $k\in\mathcal{D}$, $\mathrm{GAMP}(\mathbf{Y},\mathsf{I},\mathbf{F},\mathbf{i})$ yields the leftward flowing soft symbol beliefs

$$\mu_{\mathsf{Y}_k\to\mathsf{S}_k}(\mathsf{S})=\mathfrak{N}\left(\mathsf{Y}_k;\mathsf{S}\hat{\mathsf{H}}_k+\hat{\mathsf{I}}_k,|\mathsf{S}|^2\gamma_k^{\mathsf{H}}+\gamma_k^{\mathsf{I}}+\gamma^{(0)}\right) \tag{24}$$

that are subsequently used for decoding (as described below). Here, $\{\hat{\mathsf{H}}_k,\gamma_k^{\mathsf{H}}\}$ and $\{\hat{\mathsf{I}}_k,\gamma_k^{\mathsf{I}}\}$ play the role of soft frequency-domain channel and impulse-noise estimates, respectively.

Note that if the noise $\{i_t\}$ is *not* modeled as bursty, then there is no need to apply the forward-backward algorithm and it suffices to run $\mathrm{GAMP}(\mathbf{Y},\mathsf{I},\mathbf{F},\mathbf{i})$ only once per turbo iteration. In this case, (19) reduces to (5) and $\pi_t^{(k)}$ reduces to the time-invariant prior parameter $\pi^{(k)}$ discussed in Section II.C.

4) Symbols to Bits: The SPA dictates that the messages flowing leftward through the symbol nodes $\{\mathsf{S}_k\}$ come out unchanged, i.e., $\mu_{\mathsf{S}_k\to\mathcal{M}_k}=\mu_{\mathsf{Y}_k\to\mathsf{S}_k}$. Moreover, it dictates that the message flowing leftward out of the symbol-mapping node $\mathcal{M}_k$ and into the coded-bit node $c_{k,m}$ takes the form

$$\begin{aligned}\mu_{\mathcal{M}_k\to c_{k,m}}(c)&=\sum_{l=1}^{|\mathbb{S}|}\sum_{\mathbf{c}_k\setminus c_m}P\left(\mathsf{S}^{(l)}|\mathbf{c}_k\right)\mu_{\mathsf{S}_k\to\mathcal{M}_k}\left(\mathsf{S}^{(l)}\right) &(25)\\&\quad\times\prod_{m'\neq m}\mu_{c_{k,m'}\mathcal{M}_k\to(c_{m'})}\\&=\frac{\sum_{l:c_m^{(l)}=c}\mu_{\mathsf{S}_k\to\mathcal{M}_k}\left(\mathsf{S}^{(l)}\right)\mu_{\mathcal{M}_k\to\mathsf{S}_k}\left(\mathsf{S}^{(l)}\right)}{\mu_{c_{k,m}\to\mathcal{M}_k}(c)} &(26)\end{aligned}$$

where the last step is derived in [30].

Finally, the computed coded-bit beliefs are passed to the coding/interleaving factor node. This can be viewed as passing (extrinsic) soft information into a soft-input/soft-output (SISO) decoder, where it is treated as prior information for decoding according to the "turbo" principle. SISO decoding has been studied extensively and we refer the interested reader to [35] for a detailed account. After SISO decoding terminates, it will produce extrinsic soft information, in the form of beliefs $\{\mu_{c_{k,m}\to\mathcal{M}_k}\}$, that will be passed rightward to the symbol-mapping nodes at the start of the next turbo iteration. The turbo iterations are terminated after either the decoder detects no bit errors, the beliefs $\{\mu_{c_{k,m}\to\mathcal{M}_k}\}$ have converged, or a maximum number of turbo iterations has elapsed.

D. Simplified Receivers

Although the JCISB receiver, as presented in Section III.C, utilizes all data, pilot, and null tones to perform inference over the complete factor graph in Fig. 3, the proposed framework is flexible in that it can be easily modified to provided a desired trade-off between performance and computational complexity. For example, due to computational or architectural constraints, one might opt to simplify the receiver by either 1) using only a subset $\mathcal{U}$ of tones, or 2) replacing variable nodes in the factor graph with fixed exogenous soft estimates of those variables. (An example of a simplified receiver based on JCISB is given in [36].)

Since reducing the size of the tone subset $\mathcal{U}$ will reduce both receiver complexity and performance (see Section III.F), the selection of $\mathcal{U}$ should be done carefully to balance these conflicting objectives. In the sequel, we will denote the JCISB receiver that utilizes only the tone subset $\mathcal{U}\subset\{\mathcal{D}\cup\mathcal{P}\cup\mathcal{N}\}$ by $\mathrm{JCISB}(\mathcal{U})$. A generic implementation of $\mathrm{JCISB}(\mathcal{U})$ would execute the steps in Section III.F but with $\mathrm{GAMP}(\mathbf{Y}_{\mathcal{U}},\mathbf{H}_{\mathcal{U}},\sqrt{N}\mathbf{F}_{\mathcal{U}},\mathbf{h})$ and $\mathrm{GAMP}(\mathbf{Y}_{\mathcal{U}},\mathsf{I}_{\mathcal{U}},\mathbf{F}_{\mathcal{U}},\mathbf{i})$, and then compute approximate-MMSE estimates of $\{\mathsf{H}_k\}_{k\in\bar{\mathcal{U}}}$ and $\{\mathsf{I}_k\}_{k\in\bar{\mathcal{U}}}$ at $\bar{\mathcal{U}}=\{\mathcal{D}\cup\mathcal{P}\cup\mathcal{N}\}\setminus\mathcal{U}$ using GAMP's time-domain approximate-MMSE estimates $\{\hat{h}_t,\gamma_t^h\}$ and $\{\hat{i}_t,\gamma_t^i\}$ and the linear relationships $\mathbf{H}=\sqrt{N}\mathbf{F}\mathbf{h}$ and $\mathsf{I}=\mathbf{F}\mathbf{i}$. That said, the case $\mathcal{U}=\mathcal{P}\cup\mathcal{N}$ deserves special attention, since here it suffices to perform joint channel and impulse (JCI) estimation in a manner that is decoupled from symbol and bit estimation.

There are several ways that one might remove variable nodes from the factor graph in Fig. 3 to simplify the resulting JCISB receiver (at the expense of performance: see Section IV). For example,

1) Non-Bursty JCISB: Here the time-domain impulse-noise $\{i_t\}$ is modeled as non-bursty, in which case it suffices to remove the noise-state nodes $\{z_t\}$, use the GM prior (5) in the factor nodes $\{d_t\}$, and execute one impulse-noise iteration (without the forward-backward algorithm) per turbo iteration.

2) Joint Channel, Symbol, and Bit (JCSB) Estimation: Here we separately estimate $\{\mathsf{I}_k\}$ from only the null tones using $\mathrm{GAMP}(\mathbf{Y},\mathsf{I},\mathbf{F},\mathbf{i})$, and then fix the resulting soft estimates $\{\hat{\mathsf{I}},\gamma^{\mathsf{I}}\}$ over the turbo iterations, avoiding the need to run $\mathrm{GAMP}(\mathbf{Y},\mathsf{I},\mathbf{F},\mathbf{i})$ more than once.

3) Joint Impulse, Symbol, and Bit (JISB) Estimation: Here we compute soft linear-MMSE estimates of the frequency-domain channel coefficients $\{\mathsf{H}_k\}$ and use these in place of the GAMP-computed nonlinear-MMSE estimates $\{\hat{\mathsf{H}}_k,\gamma_k^{\mathsf{H}}\}$, avoiding the need to ever run $\mathrm{GAMP}(\mathbf{Y},\mathbf{H},\sqrt{N}\mathbf{F},\mathbf{h})$.

4) GAMP-Impulse (GI) Estimation: Here we first LMMSE estimate $\{\mathsf{H}_k\}$ from the pilot tones, then use those outputs with $\mathrm{GAMP}(\mathbf{Y},\mathsf{I},\mathbf{F},\mathbf{i})$ to estimate $\{\mathsf{I}_k\}$ from the pilot and null tones, and finally use both the soft channel and impulse estimates to recover the symbols and bits via standard SISO decoding. The principal feature distinguishing this approach from

conventional OFDM estimation is the use of GAMP-impulse estimation from pilot and null tones. The GI provides an important reference point since it uses the same information provided by the null and pilot tones as the prior work in [21]–[23].

E. Computational Complexity

The computational complexity of JCISB stems primarily from repeated calls to the GAMP algorithm, whose complexity grows as $O(N \log N + N|\mathbb{S}|)$ per GAMP iteration. For small constellations $\mathbb{S}$, GAMP's per-iteration complexity is dominated by steps (R2) and (R8) of Table I, which can each be implemented in $\frac{N}{2}\log_2 N$ multiplies using an N-length FFT; due to the constant-modulus nature of the entries of the DFT-matrix $\mathbf{\Phi}$, steps (R1) and (R7) reduce to simple summations. For large constellations like 1024-QAM, steps (R3)–(R4), which involve N_{d} summations of $|\mathbb{S}|$ terms each (see Table II), may also be of significant complexity.

As discussed, the simplifications discussed in Section III.D can be used to reduce the complexity. For example, when $\mathcal{U}$ does not include data tones, the GAMP likelihoods (16) and (18) do not involve the $|\mathbb{S}|$-term summations and so GAMP complexity is no longer dependent on $|\mathbb{S}|$. Even in that case, though, the final N_{d} symbol beliefs (24) must be evaluated at all $|\mathbb{S}|$ symbol possibilities $\mathsf{S} \in \mathbb{S}$, and so the proposed receiver complexity order remains at $O(N \log N + N|\mathbb{S}|)$, as does that of the conventional OFDM receiver.

In contrast, the state-of-the-art approach [23] uses the SBL algorithm for impulse-noise estimation, an iterative approach that computes a matrix inversion at each iteration, and thus has overall complexity $O(N^3 + N|\mathbb{S}|)$. Thus, given that N is usually in the hundreds or thousands, the proposed JCISB approach will require much less computation than [23].

F. Pilot and Null Tone Placement and Selection

In conventional OFDM systems, it is typical to place pilot tones on a uniformly spaced grid, as this yields MMSE optimal channel estimates in AWGN-corrupted frequency-selective channels [50]. Meanwhile, it is customary to place null tones at the spectrum edges in order to reduce out-of-band emissions [51]. These practices, however, should be re-examined when the receiver is expected to operate in the presence of impulsive noise, since there the MMSE channel estimator is nonlinear and the frequency-domain noise is dependent across tones, making it suboptimal to ignore null-tones while decoding.

Viewing impulse-noise estimation as a sparse reconstruction problem [21], we realize that the placement $\mathcal{U}$ of the tones used for estimation strongly affects the isometry of the linear transformation $\mathbf{F}_{\mathcal{U}}$ relating the sparse tone sequence $\mathbf{i}$ to the linearly compressed measurements $\mathbf{Y}_{\mathcal{U}}$. For sparse signal reconstruction, recovery guarantees can be stated when the measurement matrix $\mathbf{\Phi}$ has sufficiently low *coherence* [52]

$$\mu(\mathbf{\Phi}) = \max_{k,l,k\neq l} \frac{|\boldsymbol{\phi}_k^* \boldsymbol{\phi}_l|}{\|\boldsymbol{\phi}_k\|_2 \|\boldsymbol{\phi}_l\|_2} \tag{27}$$

using $\boldsymbol{\phi}_k$ to denote the kth column of $\mathbf{\Phi}$. Section IV provides evidence that $\mu(\mathbf{F}_{\mathcal{U}})$ predicts the performance of tone placement $\mathcal{U}$ in impulse-noise corrupted OFDM.

IV. Numerical Results

In this section, we evaluate the performance of our proposed JCISB receivers using Monte-Carlo simulations, comparing to both existing work and fundamental bounds. We demonstrate that, in both coded and uncoded scenarios, the proposed JCISB framework provides significant performance gains over existing techniques at a computational complexity only slightly higher than the conventional DFT receiver and thus significantly lower than the best performing prior work. In fact, we show that JCISB performs within 1 dB of theoretical performance bounds, establishing its near-optimality. Furthermore, we conduct numerical studies that investigate the impact of receiver simplifications, impulse-noise modeling and mitigation, and pilot/null tone placement.

A. Setup

Unless stated otherwise, pilot tones were spaced on a uniform grid while the null tones were placed randomly. Noise realizations were generated according to one of the two models described in Section II.C: non-bursty i.i.d-GM noise, having two impulsive noise states with powers 20 dB and 30 dB above the background noise occurring 7% and 3% or the time, respectively; and bursty GHMM noise, with the same marginal statistics but with temporal dynamics governed by the state transition matrix in (8). Unless noted otherwise, JCISB was run using at most 5 turbo iterations, 5 noise iterations, and 15 GAMP iterations. Throughout, signal-to-noise ratio (SNR) refers to the ratio of the received signal power to the total noise power.

B. Comparison With Existing Schemes

Fig. 4 plots uncoded symbol-error rate (SER) versus SNR for a prototypical PLC setting: 4-QAM modulated OFDM with 256 subcarriers, of which 80 tones are nulls and 15 are pilots, under a 5-tap Rayleigh-fading channel corrupted by i.i.d GM noise. In Fig. 4, the "JCIS" trace represents our proposed JCISB approach but without bit estimation (since here we evaluate uncoded SER), and the "GI" trace represents the proposed GI simplification from Section III.D. The "DFT" trace represents the conventional OFDM receiver, which performs LMMSE pilot-aided channel estimation, LMMSE equalization, and decoupled symbol-detection on each equalized tone. The "PP" trace refers to [13], which performs MMSE-optimal processing prior to conventional OFDM reception and has been shown to perform best among the "pre-processing" techniques discussed in Section I.B. The "SBL" trace refers to [23], which was recently shown to perform best among the "sparse reconstruction" methods. Here, the PP and SBL approaches include LMMSE channel estimation, whereas in the original formulations [13], [23] the channel was treated as known. The "MFB" trace shows the matched-filter bound, which computes tone-averaged SER assuming that each symbol is detected under perfect knowledge of every other symbol as well as the channel. By subtracting the known effect of the other symbols, the received signal under MFB is given by

$$\mathbf{y} = \mathbf{H}\mathbf{F}^*(\mathsf{S}\mathbf{e}_k) + \mathbf{n} = \mathsf{S}\bar{\mathbf{f}}_k + \mathbf{n} \tag{28}$$

where the unknown symbol S is sent on tone k and where $\mathbf{e}_k$ is the standard basis and $\bar{\mathbf{f}}_k$ is the k-th column of $\mathbf{H}\mathbf{F}^*$. Using the

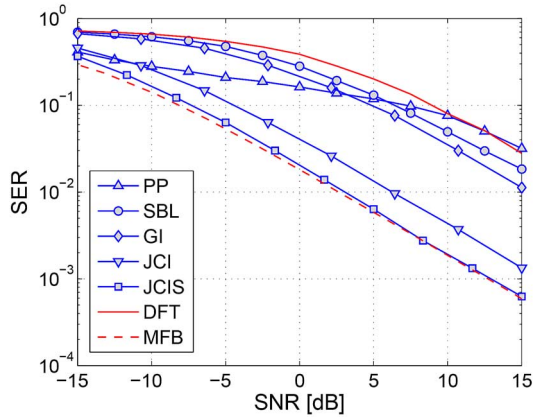


Fig. 4. Uncoded SER versus SNR for 4-QAM OFDM with 256 total tones, 80 null tones, and 15 pilot tones over a 5-tap Rayleigh-fading channel in i.i.d GM noise.

factorization of the noise pdf in time domain, it is straightforward to find the MAP detection rule for S [53]. Due to the non-Gaussianity of the noise, we evaluated the MFB via Monte-Carlo.

The SER curves in Fig. 4 show that the proposed JCIS receiver drastically outperforms the conventional OFDM receiver (by 15 dB), the PP receiver (by 15 dB in the high SNR regime), and the state-of-the-art SBL receiver (by 13 dB). We attribute these huge performance gains to the fact that JCIS utilizes *all* received tones (pilots, nulls, and data) for *joint* channel, impulse, and symbol estimation. In contrast, PP does not use OFDM signal structure for impulse-noise mitigation; and SBL decouples the estimation of the channel, impulses, and symbols, and performs linear MMSE channel estimation using only pilot tones, which not only ignores information on data and null tones, but is also strongly suboptimal in the presence of impulsive noise. Moreover, the proposed JCIS receiver follows the MF bound to within 1 dB over the full SNR range, demonstrating its near-optimality. Fig. 4 also shows that the proposed JCI simplification performs only 3 dB worse than JCIS, and that the GI simplification performs 13 dB worse than JCIS but 2 dB better than the state-of-the-art[6] SBL receiver.

C. Impact of Impulse-Noise Modeling and Mitigation

In this section, we evaluate the relative success of various strategies for modeling and mitigating impulsive noise in OFDM, again restricting our attention to uncoded transmissions. For clarity, we consider a trivial (unit-gain non-fading) channel that is perfectly known to the receiver, and thus we include no pilot tones. Without channel estimation and bit decoding, our proposed JCISB approach then reduces to JIS. Below, we compare JIS to the SBL receiver [23] and to the GI simplification proposed in Section III.D. Given the absence of pilot tones, GI and SBL are quite similar: both perform impulse-noise estimation using only null tones and in a manner that is decoupled from symbol estimation.

We first compare the noise-estimation performance of JIS, GI, and SBL using the normalized mean squared estimation

[6]Although PP outperforms both SBL and GI when $\mathsf{SNR} < 3$, the achieved SERs are unusably high.

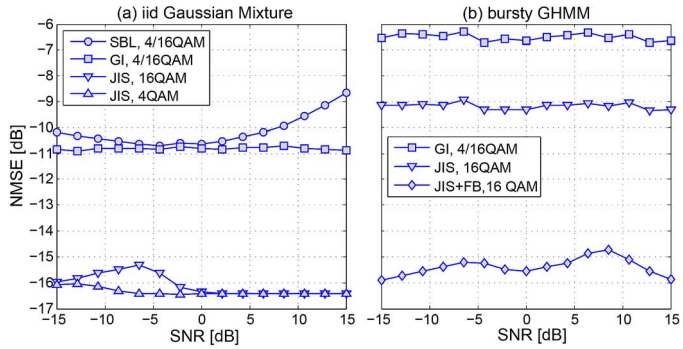


Fig. 5. NMSE versus SNR when estimating the noise sequence $\mathbf{n}$ for (a) i.i.d-GM and (b) GHMM noise models, for OFDM with 256 total tones and 60 null tones, under a known trivial channel. The corresponding SER performance is plotted in Fig. 6.

error metric $\mathsf{NMSE} = \mathbb{E}\{|n_t - \hat{i}_t|^2\}/\mathbb{E}\{|n_t|^2\} = \mathbb{E}\{|\mathsf{N}_k - \hat{\mathsf{I}}_k|^2\}/\mathbb{E}\{|\mathsf{N}_k|^2\}$, which can be interpreted as follows. Recalling that SER increases proportionally to $\mathsf{MSE} = \mathbb{E}\{|\mathsf{N}_k - \hat{\mathsf{I}}_k|^2\} = \mathbb{E}\{|\mathsf{G}_k + (\mathsf{I}_k - \hat{\mathsf{I}}_k)|^2\}$, which includes both background noise and impulse-estimation error, and noticing[7] that $\mathsf{MSE}^{-1} = \mathsf{SNR}/\mathsf{NMSE}$, we recognize NMSE as the factor by which the *effective* signal-to-noise ratio MSE^{-1} is smaller than the *stated* signal-to-noise ratio SNR.

Fig. 5(a) plots NMSE in the estimation of i.i.d GM noise versus SNR for the JIS, GI, and SBL receivers. The GI traces in Fig. 5(a) imply that GAMP is a uniformly better estimator of i.i.d GM noise than SBL, although the difference is < 1 dB for SNRs between -15 and 8 dB. This behavior is expected, given that the underlying problem is one of estimating a length-256 i.i.d-GM sequence from 60 randomly selected Fourier observations, for which the superiority of GM-GAMP over SBL was established in [54]. The JIS traces in Fig. 5(a) show NMSEs that are significantly (i.e., ≥ 5 dB) better than GI and SBL across the SNR range, and this is because JIS uses both null and data tones, rather than just null tones. To extract meaningful noise information from the data tones, JIS must accurately infer the data symbols. The latter is easier with 4-QAM than with 16-QAM, which explains the gap between the traces in Fig. 5(a).

Fig. 5(b) plots NMSE in the estimation of GHMM noise versus SNR for the proposed JIS receiver with the forward-backward (FB) iterations, and two simplifications: JIS without FB (labelled as "JIS" for consistency with Fig. 5(a)) and GI. Comparing Fig. 5(b) to (a), we see that GHMM noise is significantly more challenging than i.i.d-GM noise: the NMSE of JIS is 7 dB worse, and that of GI is 5 dB worse, in the GHMM case. However, the FB iterations help significantly: they restore approximately 6 dB of the lost NMSE.

Next we compare the SER performance of JIS, GI, and SBL in the same trivial-channel setting. Fig. 6(a) shows the case of i.i.d-GM noise. There we see that JIS significantly outperforms SBL with both 4-QAM (red) and 16-QAM (blue) constellations, as expected from the superior noise-estimation NMSE in Fig. 5 and from the fact that JIS estimates the symbols *jointly* with the noise impulses. Meanwhile, it shows GI performing on par with SBL with 4-QAM but somewhat better than SBL with 16-QAM, especially at medium SNR.

[7]For a unit signal power, $\mathsf{SNR}^{-1} = \mathbb{E}\{|\mathsf{N}_k|^2\}$, so that $\mathsf{NMSE} = \mathsf{MSE} \times \mathsf{SNR}$ and thus $\mathsf{MSE}^{-1} = \mathsf{SNR}/\mathsf{NMSE}$.

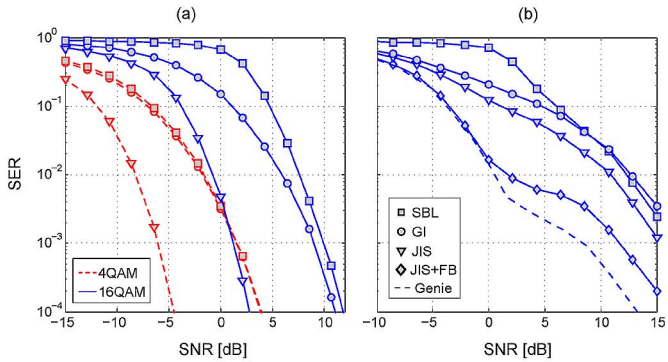


Fig. 6. Uncoded SER versus SNR for OFDM with 256 total tones and 60 null tones under a known trivial channel for (a) i.i.d GM and (b) GHMM noise. Red traces denote 4-QAM while blue traces denote 16-QAM.

Fig. 6(b) then shows SER under GHMM (i.e., bursty) noise. Comparing Fig. 6(b) to (a), we see that the burstiness of the noise causes the SER of *all* receivers to degrade significantly. Moreover, this degradation persists when the JIS receiver uses MC iterations, even though the NMSE results in Fig. 5 show only about a 1 dB loss due to burstiness. We attribute the SER sensitivity to the fact that the noise burstiness makes some OFDM-symbols much more noise-corrupted than others, and those heavily corrupted symbols lead to a degradation in SER, reported in Fig. 6(b), that exceeds the degradation expected of 1 dB i.i.d. Gaussian noise (unreported simulations show that the errors in GHMM noise estimation under JIS-FB are non-Gaussian and dependent). Regardless, Fig. 6(b) shows that the FB-assisted JIS receiver significantly outperforms non-FB-assisted JIS, GI, and the state-of-the-art SBL algorithm, especially at medium SNR. To investigate whether the kink in the JIS+FB trace was due to suboptimality of the FB noise-state inference, we simulated a genie-aided receiver that knows the true state of the GHMM noise at each time index. Since the genie trace also exhibits the kink, it is evidently not due to suboptimality of FB.

D. Impact of Pilot and Null Tone Placement

In this section, we investigate the impact of pilot and null tone placement. For this, we examine the uncoded SER of a 4-QAM 256-tone OFDM system under a 5-tap Rayleigh-fading channel in i.i.d-GM noise for both the proposed JCIS receiver and its JCI simplification, the latter of which ignores data tones during channel and impulse-noise estimation. Fig. 7 shows that the conventional placement of sideband null tones and uniform pilot tones produces the worst SER performance. Randomizing the pilot locations alone provides a modest performance gain for both JCI and JCIS, while randomizing the null locations alone improves the SER performance dramatically, especially for JCI.[8] By randomizing the locations of pilot and null tones, we effectively observe the impulsive noise vector through random projections that typically achieve good "coherence" properties (i.e., roughly the measurements have low correlation) which is beneficial for reconstruction (see [52] and references therein). We conjectured in Section III.F that the

[8]We expect JCI to be more sensitive to null/pilot-tone placement than JCIS, since the former observes the channel and noise impulses only through those tones.

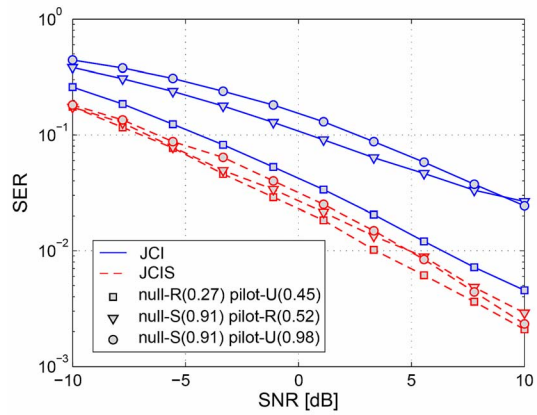


Fig. 7. Uncoded SER for OFDM with 256 total tones, 60 null tones, and 25 pilot tones under a 5-tap Rayleigh-fading channel and i.i.d-GM noise. Several null and pilot tone placements are considered: random (R), uniform (U), and sideband (S), with the corresponding coherence [$\mu(\mathbf{F}_{\mathcal{N}})$ or $\mu(\mathbf{F}_{\mathcal{P}})$, recall (27)] specified in the legend.

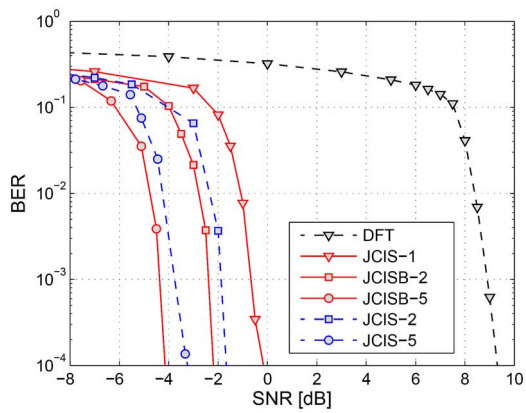


Fig. 8. Coded BER for OFDM with 1024 total tones, 150 pilot tones, and 0 null tones under a 10-tap Rayleigh-fading channel in i.i.d-GM noise. Here, JCISB-# corresponds to JCISB after # turbo iterations, and similar for JCIS-#.

performance improvement observed with randomized pilot and null tone placements can be explained by the corresponding reduction in coherence $\mu(\mathbf{F}_{\mathcal{P}})$ and $\mu(\mathbf{F}_{\mathcal{N}})$, and the coherence values reported in Fig. 7 lend credence to this conjecture.

E. Coded Systems

Finally, we investigate the bit error rate (BER) performance of JCISB in the coded scenario. For this, we used an LDPC-coded 16-QAM 1024-tone OFDM system with 150 pilot and 0 null tones under a 10-tap Rayleigh-fading channel and i.i.d-GM noise. The LDPC codes had code-word length ≈ 60000 and rate 1/2, with a modified coder/decoder implementations from [55]. We also investigate the conventional OFDM receiver ("DFT") as well as the JCIS simplification, which omits SISO decoding from the turbo iterations. For both JCIS and DFT, we performed SISO decoding as the final step. For all receivers, the maximum number of LDPC iterations was 50.

Fig. 8 shows that, after only one turbo iteration, the proposed JCISB[9] outperforms the conventional OFDM receiver by 10 dB. Additional turbo iterations result in further gains of 4 dB. Fig. 8 also shows that JCIS's decoupling of bit estimation from

[9]Note that, with only a single turbo iteration, JCISB and JCIS are equivalent.

channel, impulse, and symbol estimation costs approximately 1 dB.

V. Conclusion

In a this paper, we presented a factor-graph approach to OFDM reception in multipath distorted and impulse-noise corrupted channels that performs near-optimal joint channel, impulse-noise, symbol, and bit (JCISB) estimation. Our approach merges recent work on generalized approximate message passing (GAMP) [31], its "turbo" extension to larger factor graphs [34], and soft-input-soft-output SISO decoding [35]. Extensive numerical simulations show that the proposed JCISB receiver provides drastic performance gains over existing receivers for OFDM in impulsive noise, and performs within 1dB of the matched-filter bound, all while matching the complexity order of the conventional OFDM receiver. Furthermore, JCISB is easily parallelized, providing a natural mapping to FPGA implementations (see [36] for a recent FPGA implementation of the GI receiver). Additional numerical simulations investigated the impact of JCISB simplifications, noise modeling and mitigation, and null/pilot tone placement. Throughout this paper we have assumed prior knowledge of the noise parameters while in practical settings this information might have to be estimated by the receiver. Incorporating parameter estimation in the proposed framework is an interesting topic for future work.

Appendix A
Generalized Approximate Message Passing (GAMP)

The GAMP algorithm, an extension of AMP from [46], was developed in [31] to address the estimation of a vector of independent possibly-non-Gaussian random variables $\mathbf{x}$ that are linearly mixed via a linear transform $\mathbf{\Phi} \in \mathbb{C}^{M\times N}$ to form $\mathbf{z} = \mathbf{\Phi x} = [z_1 \cdots z_M]^T$ and subsequently observed as $\mathbf{y} = [y_1 \cdots y_M]^T$ according to the general likelihood function $p(\mathbf{y}|\mathbf{z}) = \prod_{i=1}^{M} p(y_i|z_i)$. The GAMP algorithm is intended for the case when the dimensions M and N are both large, in which case the central limit theorem suggests approximating the product of messages flowing leftward into each factor node $f_i = p(y_i|z_i)$ in Fig. 9 as $\prod_j \mu_{x_j\to f_i}(z_i) \approx \mathfrak{N}(z_i;\hat{p}_i,\gamma_i^p)$ and the product of messages flowing rightward into each variable node x_j as $\prod_i \mu_{f_i\to x_j}(x_j) \approx \mathfrak{N}(x_j;\hat{r}_j,\gamma_j^r)$, where the quantities $\hat{p}_i$, γ_i^p, $\hat{r}_j$, and γ_j^r can be computed from Table I. Similarly, to remain computationally tractable, the pdf representing each message leaving a factor or a variable node is simplified by taking its second order Taylor series expansion. As a result, each of those messages can be approximated by two parameters resulting from the Taylor expansion. The corresponding "GAMP$(\mathbf{y},\mathbf{z},\mathbf{\Phi},\mathbf{x})$" algorithm is summarized in Table I. The detailed derivation and theoretical guarantees of the GAMP algorithm are beyond the scope of this paper; we refer the interested reader to [31] and [48] for more information.

Appendix B
Derivation of GAMP$(\mathbf{Y},\mathbf{H},\sqrt{N}\mathbf{F},\mathbf{h})$ Likelihood

This Appendix derives the output MMSE estimation functions $\mathbb{E}\{\mathsf{H}_k|\mathbf{Y};\hat{p},\gamma^p\}$ and $\mathbb{V}\{\mathsf{H}_k|\mathbf{Y};\hat{p},\gamma^p\}$ used in steps

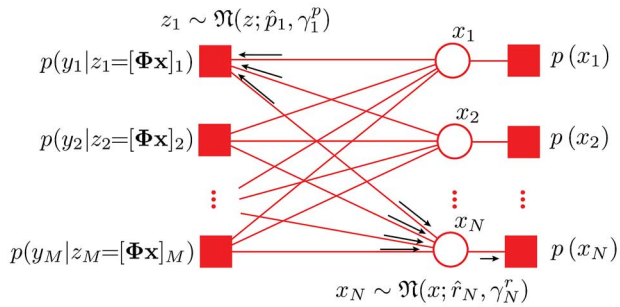


Fig. 9. Factor graph used to derive GAMP with illustrations of message approximations.

(R3)–(R4) of GAMP$(\mathbf{Y},\mathbf{H},\sqrt{N}\mathbf{F},\mathbf{h})$. We start with the case of data tones $k \in \mathcal{D}$, where

$$\mathsf{Y}_k = \mathsf{S}_k\mathsf{H}_k + \mathsf{I}_k + \mathsf{G}_k.$$

Y_k is Gaussian when conditioned on S_k, and so according to the definition (D1) in Table I this is a standard linear MMSE problem whose solution is given by ([56], Ch. 12),

$$\mathbb{E}\{\mathsf{H}_k|\mathsf{S}_k,\mathsf{Y}_k;\hat{p},\gamma^p\} = \hat{p} + \frac{\mathsf{S}_k^*\gamma^p(\mathsf{Y}_k - \hat{\mathsf{I}}_k - \mathsf{S}_k\hat{p})}{\gamma^{(0)} + \gamma_k^{\mathsf{I}} + |\mathsf{S}_k|^2\gamma^p}. \quad (29)$$

Given the belief $\{\beta_k^{(l)}\}_{l=1}^{|\mathbb{S}|}$ about symbol S_k and the decoupling achieved by GAMP between H_k and $\mathbf{Y}$ (H_k depends only on Y_k), the *law of total expectation* implies

$$\mathbb{E}\{\mathsf{H}_k|\mathbf{Y};\hat{p},\gamma^p\} = \mathbb{E}_{\mathsf{S}_k|\mathsf{Y}_k}\{\mathbb{E}\{\mathsf{H}_k|\mathsf{S}_k,\mathsf{Y}_k;\hat{p},\gamma^p\}\} \quad (30)$$

$$= \hat{p} + \sum_{l=1}^{|\mathbb{S}|}\lambda_k^{(l)}\frac{\gamma^{\mathsf{I}}\left(\mathsf{Y}_k - \hat{\mathsf{I}}_k - \hat{p}\mathsf{S}^{(l)}\right)}{\gamma^{(0)} + \gamma_k^{\mathsf{I}} + |\mathsf{S}^{(l)}|^2\gamma^p} \quad (31)$$

where

$$\lambda_k^{(l)} = P\left(\mathsf{S}_k = \mathsf{S}^{(l)}|\mathsf{Y}_k;\hat{p},\gamma^p\right) \propto p(\mathsf{Y}_k|\mathsf{S}_k;\hat{p},\gamma^p)\beta_k^{(l)} \quad (32)$$

is the posterior symbol probability and $p(\mathsf{Y}_k|\mathsf{S}_k;\hat{p},\gamma^p) = \mathfrak{N}(\mathsf{Y}_k;\hat{\mathsf{I}}_k + \hat{p}\mathsf{S}_k,\gamma^{\mathsf{I}} + |\mathsf{S}_k|^2\gamma^p + \gamma^{(0)})$. Similarly, the *law of total variance* implies

$$\mathbb{V}\{\mathsf{H}_k|\mathbf{Y};\hat{p},\gamma^p\}$$
$$= \mathbb{E}_{\mathsf{S}_k|\mathsf{Y}_k}\{\mathbb{V}\{\mathsf{H}_k|\mathsf{S}_k,\mathsf{Y}_k;\hat{p},\gamma^p\}\} + \mathbb{V}_{\mathsf{S}_k|\mathsf{Y}_k}\{\mathbb{E}\{\mathsf{H}_k|\mathsf{S}_k,\mathsf{Y}_k;\hat{p},\gamma^p\}\} \quad (33)$$

$$= \sum_{l=1}^{|\mathbb{S}|}\lambda_k^{(l)}\left[\frac{\gamma^p\left(\gamma^{(0)} + \gamma_k^{\mathsf{I}}\right)}{\gamma^{(0)} + \gamma_k^{\mathsf{I}} + |\mathsf{S}^{(l)}|^2\gamma^p} + |\mathbb{E}\{\mathsf{H}_k|\mathsf{S}_k,\mathsf{Y}_k;\hat{p},\gamma^p\}|^2\right] - |\mathbb{E}\{\mathsf{H}_k|\mathsf{Y}_k;\hat{p},\gamma^p\}|^2. \quad (34)$$

The derivation for pilot tones $k \in \mathcal{P}$ reduces to the above under $\lambda_k^{(1)} = 1$, $\lambda_k^{(l\neq 1)} = 0$, and $\mathsf{S}^{(1)} = \mathsf{p}$.

Appendix C
Derivation of GAMP$(\mathbf{Y},\mathbf{I},\mathbf{F},\mathbf{i})$ Likelihood

This Appendix derives the output MMSE estimation functions $\mathbb{E}\{\mathsf{I}_k|\mathbf{Y};\hat{p},\gamma^p\}$ and $\mathbb{V}\{\mathsf{I}_k|\mathbf{Y};\hat{p},\gamma^p\}$ used in steps

(R9)–(R10) of $\mathrm{GAMP}(\mathbf{Y}, \mathbf{I}, \mathbf{F}, \mathbf{i})$. We start with the case of data tones $k \in \mathcal{D}$, where

$$\mathsf{Y}_k = \mathsf{I}_k + \mathsf{S}_k\mathsf{H}_k + \mathsf{G}_k.$$

Y_k is Gaussian when conditioned on S_k, and so according to the definition (D1) in Table I this is a standard linear MMSE problem whose solution is given by ([56], Ch. 12),

$$\mathbb{E}\{\mathsf{I}_k|\mathsf{S}_k, \mathsf{Y}_k; \hat{p}, \gamma^p\} = \hat{p} + \frac{\gamma^p(\mathsf{Y}_k - \hat{p} - \hat{\mathsf{H}}_k\mathsf{S}_k)}{\gamma^{(0)} + \gamma^p + |\mathsf{S}_k|^2\gamma_k^{\mathsf{H}}}. \quad (35)$$

Given the belief $\{\boldsymbol{\beta}_k^{(l)}\}_{l=1}^{|\mathbb{S}|}$ about symbol S_k and the decoupling achieved by GAMP between I_k and $\mathbf{Y}$ (I_k depends only on Y_k), the *law of total expectation* implies

$$\mathbb{E}\{\mathsf{I}_k|\mathbf{Y}; \hat{p}, \gamma^p\} = \mathbb{E}_{\mathsf{S}_k|\mathsf{Y}_k}\{\mathbb{E}\{\mathsf{I}_k|\mathsf{S}_k, \mathsf{Y}_k; \hat{p}, \gamma^p\}\} \quad (36)$$

$$= \hat{p} + \sum_{l=1}^{|\mathbb{S}|} \lambda_k^{(l)} \frac{\gamma^p\left(\mathsf{Y}_k - \hat{p} - \hat{\mathsf{H}}_k\mathsf{S}^{(l)}\right)}{\gamma^{(0)} + \gamma^p + |\mathsf{S}^{(l)}|^2\gamma_k^{\mathsf{H}}} \quad (37)$$

where $\lambda_k^{(l)}$ is the posterior symbol probability from (32) but now with $p(\mathsf{Y}_k|\mathsf{S}_k; \hat{p}, \gamma^p) = \mathfrak{N}(\mathsf{Y}_k; \hat{p} + \hat{\mathsf{H}}_k\mathsf{S}_k, \gamma^p + |\mathsf{S}_k|^2\gamma_k^{\mathsf{H}} + \gamma^{(0)})$. Similarly, the *law of total variance* implies

$$\begin{aligned}
&\mathbb{V}\{\mathsf{I}_k|\mathsf{Y}_k; \hat{p}, \gamma^p\} \\
&\quad = \mathbb{E}_{\mathsf{S}_k|\mathsf{Y}_k}\{\,\mathbb{V}\{\mathsf{I}_k|\mathsf{S}_k, \mathsf{Y}_k; \hat{p}, \gamma^p\}\} \\
&\qquad + \mathbb{V}_{\mathsf{S}_k|\mathsf{Y}_k}\{\mathbb{E}\{\mathsf{I}_k|\mathsf{S}_k, \mathsf{Y}_k; \hat{p}, \gamma^p\}\} \\
&\quad = \sum_{l=1}^{|\mathbb{S}|} \lambda_k^{(l)} \left[\frac{\gamma^p\left(\gamma^{(0)} + |\mathsf{S}^{(l)}|^2\gamma_k^{\mathsf{H}}\right)}{\gamma^{(0)} + \gamma^p + |\mathsf{S}^{(l)}|^2\gamma_k^{\mathsf{H}}}\right. \\
&\qquad \left. + |\mathbb{E}\{\mathsf{I}_k|\mathsf{S}_k, \mathsf{Y}_k; \hat{p}, \gamma^p\}|^2\right] \\
&\qquad - |\mathbb{E}\{\mathsf{I}_k|\mathsf{Y}_k; \hat{p}, \gamma^p\}|^2. \quad (38)
\end{aligned}$$

The derivation for pilot tones $k \in \mathcal{P}$ reduces to the above under $\lambda_k^{(1)} = 1$, $\lambda_k^{(l\neq 1)} = 0$, and $\mathsf{S}^{(1)} = \mathsf{p}$. Meanwhile, the derivation for null tones $k \in \mathcal{N}$ is the special case of pilots with $\mathsf{p} = 0$.

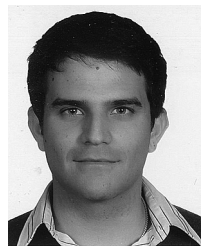
ACKNOWLEDGMENT

The authors would like to express gratitude to Dr. Omar El Ayach for many useful discussions and suggestions.

REFERENCES

[1] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[2] K. Blackard, T. Rappaport, and C. Bostian, "Measurements and models of radio frequency impulsive noise for indoor wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 11, no. 7, pp. 991–1001, 1993.
[3] W. Lauber and J. Bertrand, "Statistics of motor vehicle ignition noise at VHF/UHF," *IEEE Trans. Electromagn. Compat.*, vol. 41, no. 3, pp. 257–259, 1999.
[4] M. Sanchez, L. de Haro, M. Ramon, A. Mansilla, C. Ortega, and D. Oliver, "Impulsive noise measurements and characterization in a UHF digital TV channel," *IEEE Trans. Electromagn. Compat.*, vol. 41, no. 2, pp. 124–136, 1999.
[5] M. Sanchez, A. Alejos, and I. Cuinas, "Urban wide-band measurement of the UMTS electromagnetic environment," *IEEE Trans. Veh. Technol.*, vol. 53, no. 4, pp. 1014–1022, 2004.
[6] M. Nassar, X. Lin, and B. L. Evans, "Stochastic modeling of microwave oven interference in WLANs," in *Proc. IEEE Int. Conf. Commun.*, 2011, pp. 1–6.
[7] M. Nassar, K. Gulati, M. DeYoung, B. L. Evans, and K. Tinsley, "Mitigating near-field interference in laptop embedded wireless transceivers," *J. Signal Process. Syst.*, vol. 63, pp. 1–12, 2011.
[8] M. Zimmermann and K. Dostert, "Analysis and modeling of impulsive noise in broad-band powerline communications," *IEEE Trans. Electromagn. Compat.*, vol. 44, no. 1, pp. 249–258, 2002.
[9] M. Nassar, K. Gulati, Y. Mortazavi, and B. L. Evans, "Statistical modeling of asynchronous impulsive noise in powerline communication networks," in *Proc. IEEE Global Commun. Conf.*, 2011, pp. 1–6.
[10] M. Nassar, J. Lin, Y. Mortazavi, A. Dabak, I. H. Kim, and B. L. Evans, "Local utility power line communications in the 3–500 KHz band: Channel impairments, noise, and standards," *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 116–127, 2012.
[11] S. Zhidkov, "Analysis and comparison of several simple impulsive noise mitigation schemes for OFDM receivers," *IEEE Trans. Commun.*, vol. 56, no. 1, pp. 5–9, 2008.
[12] D.-F. Tseng, Y. Han, W. H. Mow, L.-C. Chang, and A. Vinck, "Robust clipping for OFDM transmissions over memoryless impulsive noise channels," *IEEE Commun. Lett.*, vol. 16, no. 7, pp. 1110–1113, 2012.
[13] J. Haring, *Error Tolerant Communication Over the Compound Channel*. New York, NY, USA: Springer-Verlag, 2002.
[14] J. Haring and A. Vinck, "Iterative decoding of codes over complex numbers for impulsive noise channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1251–1260, 2003.
[15] A. Mengi and A. Vinck, "Successive impulsive noise suppression in OFDM," in *Proc. IEEE Int. Symp. Power Line Commun. Appl.*, 2010, pp. 33–37.
[16] M. Nassar and B. L. Evans, "Low complexity EM-based decoding for OFDM systems with impulsive noise," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, 2011, pp. 1943–1947.
[17] C.-H. Yih, "Iterative interference cancellation for OFDM signals with blanking nonlinearity in impulsive noise channels," *IEEE Signal Process. Lett.*, vol. 19, no. 3, pp. 147–150, 2012.
[18] J. Wolf, "Redundancy, the discrete Fourier transform, and impulse noise cancellation," *IEEE Trans. Commun.*, vol. COM-31, no. 3, pp. 458–461, 1983.
[19] F. Abdelkefi, P. Duhamel, and F. Alberge, "Impulsive noise cancellation in multicarrier transmission," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 94–106, 2005.
[20] F. Abdelkefi, P. Duhamel, and F. Alberge, "A necessary condition on the location of pilot tones for maximizing the correction capacity in OFDM systems," *IEEE Trans. Commun.*, vol. 55, no. 2, pp. 356–366, 2007.
[21] G. Caire, T. Al-Naffouri, and A. Narayanan, "Impulse noise cancellation in OFDM: An application of compressed sensing," in *Proc. IEEE Int. Symp. Inf. Theory*, 2008, pp. 1293–1297.
[22] L. Lampe, "Bursty impulse noise detection by compressed sensing," in *Proc. IEEE Int. Symp. Power Line Commun. Appl.*, 2011, pp. 29–34.
[23] J. Lin, M. Nassar, and B. L. Evans, "Impulsive noise mitigation in powerline communications using sparse Bayesian learning," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1172–1183, Jul. 2013.
[24] A. Mehboob, L. Zhang, J. Khangosstar, and K. Suwunnapuk, "Joint channel and impulsive noise estimation for OFDM based power line communication systems using compressed sensing," in *Proc. IEEE Int. Symp. Power Line Commun. Appl.*, 2013, pp. 203–208.
[25] A. P. Worthen and W. Stark, "Unified design of iterative receivers using factor graphs," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 843–849, 2001.
[26] C. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.
[27] 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.*, 2009, pp. 2561–2564.
[28] Y. Liu, L. Brunel, and J. Boutros, "Joint channel estimation and decoding using Gaussian approximation in a factor graph over multipath channel," in *Proc. IEEE Int. Symp. Pers., Indoor Mobile Radio Commun.*, 2009, pp. 3164–3168.
[29] G. Kirkelund, C. Manchon, L. Christensen, E. Riegler, and B. Fleury, "Variational message-passing for joint channel estimation and decoding in MIMO-OFDM," in *Proc. IEEE Global Telecommun. Conf.*, 2010, pp. 1–6.
[30] P. Schniter, "A message-passing receiver for BICM-OFDM over unknown clustered-sparse channels," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 8, pp. 1462–1474, 2011.
[31] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inf. Theory*, 2011, pp. 2174–2178, (See also longer version in Arxiv:1010.5141).
[32] P. Schniter, "Belief-propagation-based joint channel estimation and decoding for spectrally efficient communication over unknown sparse channels," *Phys. Commun.*, vol. 5, no. 3, pp. 91–101, 2012.
[33] A. P. Kannu and P. Schniter, "On communication over unknown sparse frequency-selective block-fading channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 6619–6632, Oct. 2011.

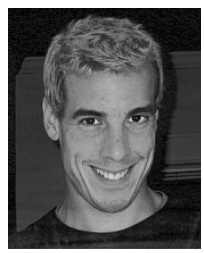
[34] P. Schniter, "Turbo reconstruction of structured sparse signals," in *Proc. Conf. Inf. Sci. Syst.*, 2010, pp. 1–6.
[35] D. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge, MA, USA: Cambridge Univ. Press, 2003.
[36] K. F. Nieman, M. Nassar, J. Lin, and B. L. Evans, "FPGA implementation of a message-passing OFDM receiver for impulsive noise channels," presented at the Asilomar Conf. Signals, Syst., Comput., Pacific Grove, CA, USA, Nov. 3–6, 2013.
[37] P. Schniter and D. Meng, "A message-passing receiver for BICM-OFDM over unknown time-varying sparse channels," presented at the Allerton Conf. Commun., Control, Comput., Monticello, IL, USA, Sep. 2011, Invited Paper.
[38] D. Middleton, "Non-Gaussian noise models in signal processing for telecommunications: New methods an results for class A and class B noise models," *IEEE Trans. Inf. Theory*, vol. 45, no. 4, pp. 1129–1149, 1999.
[39] J. Ilow and D. Hatzinakos, "Analytic alpha-stable noise modeling in a Poisson field of interferers or scatterers," *IEEE Trans. Signal Process.*, vol. 46, no. 6, pp. 1601–1611, 1998.
[40] K. Gulati, B. L. Evans, J. Andrews, and K. Tinsley, "Statistics of co-channel interference in a field of Poisson and Poisson-Poisson clustered interferers," *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 6207–6222, 2010.
[41] S. Fruhwirth-Schnatter, *Finite Mixture and Markov Switching Models*. New York, NY, USA: Springer, 2006.
[42] G. F. Cooper, "The computational complexity of probabilistic inference using Bayesian belief networks," *Artif. Intell.*, vol. 42, pp. 393–405, 1990.
[43] J. Boutros and G. Caire, "Iterative multiuser joint decoding: Unified framework and asymptotic analysis," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 1772–1793, 2002.
[44] D. Guo and C.-C. Wang, "Random sparse linear systems observed via arbitrary channels: A decoupling principle," in *Proc. IEEE Int. Symp. Inf. Theory*, 2007, pp. 946–950.
[45] R. McEliece, D. MacKay, and J.-F. Cheng, "Turbo decoding as an instance of Pearl's belief propagation algorithm," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 2, pp. 140–152, 1998.
[46] D. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," in *Proc. Natl. Acad. Sci.*, 2009, vol. 106, pp. 18 914–18 919.
[47] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 764–785, 2011.
[48] A. Javanmard and A. Montanari, "State evolution for general approximate message passing algorithms, with applications to spatial coupling," Nov. 2012, arXiv:1211.5164.
[49] H. Monajemi, S. Jafarpour, and M. Gavish, "Stat 330/CME 362 collaboration, and D. L. Donoho Deterministic matrices matching the compressed sensing phase transitions of Gaussian random matrices," in *Proc. Natl. Acad. Sci.*, Jan. 2013, vol. 110, no. 4, pp. 1181–1186.
[50] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," *IEEE Trans. Consum. Eletron.*, vol. 44, pp. 1122–1128, Aug. 1998.
[51] Powerline Related Intelligent Metering Evolution (PRIME), Prime Alliance Std. [Online]. Available: http://www.prime-alliance.org
[52] C. Studer, P. Kuppinger, G. Pope, and H. Bolcskei, "Recovery of sparsely corrupted signals," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3115–3130, 2012.
[53] A. Spaulding and D. Middleton, "Optimum reception in an impulsive interference environment-part I: Coherent detection," *IEEE Trans. Commun.*, vol. COM-25, no. 9, pp. 910–923, 1977.
[54] J. P. Vila and P. Schniter, "Expectation-maximization gaussian-mixture approximate message passing," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4658–4672, Oct. 2013.
[55] I. Kozintsev, Matlab Programs for Encoding and Decoding of LDPC Codes Over GF(2M) [Online]. Available: http://www.kozintsev.net/soft.html
[56] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.



Marcel Nassar (M'05) received the B.E. degree in computer and communications engineering, with a minor degree in mathematics, from the American University of Beirut in 2006, and the M.S. and Ph.D. degrees in electrical and computer engineering from The University of Texas at Austin in 2008 and 2013, respectively.

During his studies, he interned at the University of California, Berkeley in 2005, at Intel Corporation in 2008 and 2009, and at Texas Instruments in 2011 and 2012. After receiving his Ph.D. degree, he joined the Mobile Solutions Lab at Samsung Research America, CA as a Senior Research Engineer. His interests lie at the intersection of signal processing and machine learning with applications to statistical modeling and mitigation of interference in communication systems.

Dr. Nassar received the Best Paper Award at the 2013 IEEE International Symposium on Power Line Communications and Its Applications, and second place in the Student Paper Contest at the 2013 Asilomar Conference on Signals, Systems, and Computers.



Philip Schniter (F'14) received the B.S. and M.S. degrees in electrical engineering from the University of Illinois at Urbana-Champaign in 1992 and 1993, respectively, and the Ph.D. degree in electrical engineering from Cornell University, Ithaca, NY, in 2000.

From 1993 to 1996, he was with Tektronix Inc., Beaverton, OR, as a systems engineer. After receiving the Ph.D. degree, he joined the Department of Electrical and Computer Engineering at The Ohio State University, Columbus, where he is currently a Professor and a member of the Information Processing Systems (IPS) Lab. In 2008–2009, he was a Visiting Professor at Eurecom, Sophia Antipolis, France, and Supélec, Gif-sur-Yvette, France. His areas of interest currently include signal processing, machine learning, and wireless communications and networks.

Dr. Schniter received the National Science Foundation CAREER Award in 2003, he was elected to serve on the IEEE SPCOM Technical Committee from 2005 to 2010 and on the IEEE SAM Technical Committee since 2013.

Brian L. Evans (F'09) received the B.S.E.E.C.S (1987) degree from the Rose-Hulman Institute of Technology, and the MSEE (1988) and Ph.D. E.E. (1993) degrees from the Georgia Institute of Technology. From 1993 to 1996, he was a Postdoctoral Researcher at the University of California, Berkeley. In 1996, he joined the faculty at UT Austin.

He holds the Engineering Foundation Professorship at UT Austin. His research bridges the gap between digital signal processing theory and embedded real-time implementation. His research interests include wireless interference mitigation, smart grid communications, smart phone video acquisition, and cloud radio access networks. Prof. Evans has published more than 220 refereed conference and journal papers, and graduated 21 Ph.D. and 9 M.S. students.

Dr. Evans received the Best Paper Award at the 2013 IEEE International Symposium on Power Line Communications and Its Applications, and a Top 10% Paper Award at the 2012 IEEE Multimedia Signal Processing Workshop. He has also received three teaching awards at UT Austin: Gordon Lepley Memorial ECE Teaching Award (2008), Texas Exes Teaching Award (2011) and HKN Outstanding ECE Professor (2012). He received a 1997 US National Science Foundation CAREER Award.