Joint Channel-Estimation/Decoding With Frequency-Selective Channels and Few-Bit ADCs

Peng Sun, Zhongyong Wang, Robert W. Heath, Jr., *Fellow, IEEE*, and Philip Schniter, *Fellow, IEEE*

***Abstract*—We propose a fast and near-optimal approach to joint channel-estimation, equalization, and decoding of coded single-carrier (SC) transmissions over frequency-selective channels with few-bit analog-to-digital converters (ADCs). Our approach leverages parametric bilinear generalized approximate message passing to reduce the implementation complexity of joint channel estimation and (soft) symbol decoding to that of a few fast Fourier transforms. Furthermore, it learns and exploits sparsity in the channel impulse response. This paper is motivated by millimeter-wave systems with bandwidths on the order of Gsamples/sec, where few-bit ADCs, SC transmissions, and fast processing all lead to significant reductions in power consumption and implementation cost. We numerically demonstrate our approach using signals and channels generated according to the IEEE 802.11ad wireless local area network standard, in the case that the receiver uses analog beamforming and a single ADC.**

***Index Terms*—Low resolution analog-to-digital converter, millimeter wave, joint channel estimation and equalization, turbo equalization, approximate message passage.**

I. Introduction

THE trend towards ever-wider-bandwidths in communications systems results in major implementational challenges. This trend is evident in millimeter-wave (mmWave) systems, which exploit large chunks of bandwidth at carrier frequencies of 30 GHz and above [1]. For example, the IEEE 802.11ad standard [2] specifies channels of bandwidth 1.76 GHz centered near 60 GHz. Future 5G cellular systems are also likely to incorporate mmWave technology [3], [4].

Manuscript received February 26, 2018; revised July 6, 2018, October 15, 2018, and December 8, 2018; accepted December 8, 2018. Date of publication December 24, 2018; date of current version January 4, 2019. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Youngchul Sung. This work was supported in part by the National Science Foundation under Grants CCF-1527079 and CCF-1527162, and in part by the National Natural Science Foundation of China under Grant NSFC-61571402. This paper was presented in part at the Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, Oct./Nov. 2017. *(Corresponding author: Philip Schniter.)*

P. Sun is with the School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China. He is also with the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210 USA (e-mail: sun.1771@osu.edu).

Z. Wang is with the School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China (e-mail: iezywang@zzu.edu.cn).

R. W. Heather, Jr., is with the Wireless Networking and Communications Group, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: rheath@utexas.edu).

P. Schniter is with the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210 USA (e-mail: schniter.1@osu.edu).

Digital Object Identifier 10.1109/TSP.2018.2887189

A main challenge in wideband systems comes from the analog-to-digital converters (ADCs) used at the receiver. At bandwidths above 1 Gs/sec, ADC power consumption grows approximately quadratically with bandwidth [5], [6]. Meanwhile, ADC power consumption grows exponentially in the number of bits used in conversion. At GHz bandwidths, many-bit (e.g., 10 bit) ADCs may consume several watts of power, which is impractical for handheld mobile devices. For this reason, there has been a growing interest in few-bit (i.e., 1–4 bit) ADCs for communications receivers (e.g., [7]–[25]).

Wide bandwidth also results in challenges at the transmitter. In particular, wide-bandwidth linear amplifiers are expensive in terms of power consumption and cost [26]. For this reason, it is beneficial to transmit signals with low peak-to-average power ratio (PAPR), which allow power-amplifier linearity requirements to be relaxed. The desire for low PAPR suggests single-carrier (SC) transmission, as opposed to multi-carrier transmission such as orthogonal frequency division multiplexing (OFDM) [27]. Because wide bandwidth receivers may need to decode billions of bits per second, it is important that the SC transmission is amenable to computationally efficient channel-equalization, e.g., via fast Fourier transform (FFT) processing [26].

Although wide bandwidth brings many challenges, there is a silver lining: the measured channel responses are relatively sparse in the angle and delay domains, in both indoor [28] and outdoor [29], [30] settings. With sparse channels, the fundamental performance of a communications link can be significantly improved (e.g., [31], [32]).

We now review relevant existing work on few-bit-ADC receiver design. For flat-fading multiple-input/multiple-output (MIMO) channels, channel estimation (e.g., [7]–[11]), symbol detection (e.g., [12]–[16]), and joint channel estimation and symbol detection (e.g., [17], [18]) have been considered. However, wideband channels are frequency selective in practice.

For frequency-selective channels, channel estimation has been considered in [19], [20] using comb-type pilots that allow the channel to be treated as effectively flat-fading, but these approaches perform poorly under PAPR limits. Channel estimation for 2-tap channels was considered in [21], but realistic wideband channels have many more taps. An approach for longer channels was recently proposed in [22], but it applies only to OFDM. An iterative expectation-maximization (EM)-like channel estimation scheme for SC transmissions was proposed in [23], but it is computationally expensive and does not leverage sparsity. More recently, pilot-aided sparsity-exploiting channel-

estimation schemes were proposed in [24], and a known-channel symbol-detection scheme was proposed in [25]. Both [24] and [25] are made computationally efficient by the use of generalized approximate message passing (GAMP) [33] and FFT processing. But, as we will show, significantly improved performance can be obtained through *joint* channel estimation, symbol detection, and bit decoding. A joint channel-estimation/decoding approach was proposed in [34], but it does not leverage sparsity and requires OFDM.

In this paper, we propose a computationally efficient approach to joint channel-estimation, equalization, and decoding of single-carrier transmissions over frequency-selective channels with few-bit ADCs. Our approach is an instance of turbo-equalization [35], [36], which iterates soft equalization (and, in our case, joint channel estimation) with soft decoding. For joint channel estimation and equalization, we use the recently proposed Parametric Bilinear GAMP (PBiGAMP) framework [37], which—when specialized to our application—consumes only a few FFTs per equalizer iteration and demands relatively few equalizer iterations. We then mate PBiGAMP to the soft decoder using the turbo-AMP framework from [38]. To exploit the channel's (approximate) sparsity, we use a Gaussian mixture model (GMM), as in [39], and learn the GMM parameters via the EM algorithm, building on [40]. Portions of this work were published in [41]. Relative to [41], this paper includes detailed derivations and explanations, a refined channel-estimation scheme, and additional numerical experiments.

In this work, we assume the use of analog beamforming, and thus a single (few-bit) ADC, at the receiver. Our approach can be contrasted with digital (e.g., [24]) or hybrid (e.g., [42]) beamforming, which requires the use of multiple ADCs. It is possible that, for large arrays, with our architecture, the power consumption of the analog beamforming becomes more significant than that of the ADCs; The exact calculation is architecture-specific (see, e.g., [43]) and we leave an investigation of these issues to future work. Extensions of our approach to digital beamforming systems and to hybrid analog/digital systems are worthwhile, but outside the scope of this work. To evaluate our receiver design, we consider a system that complies with the IEEE 802.11ad 60 GHz mmWave standard [2], which supports analog beamforming. Our numerical results for the IEEE 802.11ad "conference room" channel [44] (under perfect synchronization) show only a 3 dB SNR gap at a BER of 10^{-2} for a 2-bit ADC compared to infinite bit resolution also using joint decoding. Further, we show how embracing the nonlinearity of the quantization helps to avoid a substantial SNR gap that arises when pilot-only channel estimation is used or when Bussgang linearization is used with very-few-bit ADCs at high SNR.

The paper is organized as follows. In Section II, we present our models for SC block transmission, channel propagation, and few-bit reception, as well the GMM-based channel model that we use with PBiGAMP. In Section III, after a brief introduction to belief propagation and PBiGAMP, we propose our soft joint channel-estimation/decoding method and describe how it can be mated with a soft decoder. We also describe our EM-based method to learn the GMM channel parameters. In Section IV, we detail several benchmarks that will be used in our numerical comparisons, including Bussgang-linearized PBiGAMP and linear-MMSE symbol decoding with pilot-aided channel estimation. In Section V, we report numerical results, and in Section VI we conclude.

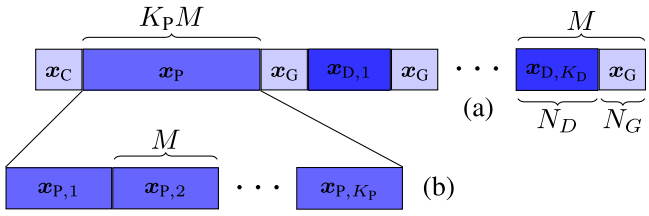


Fig. 1. (a) The transmission structure, containing cyclic-prefixed pilots $[\boldsymbol{x}_{\mathrm{C}}, \boldsymbol{x}_{\mathrm{P}}]$ and data blocks $\boldsymbol{x}_{\mathrm{D},k}$ separated by guard blocks $\boldsymbol{x}_{\mathrm{G}}$. (b) The block structure of the pilot sequence $\boldsymbol{x}_{\mathrm{P}}$.

Notation—We use boldface uppercase letters like $\boldsymbol{B}$ to denote matrices and boldface lowercase letters like $\boldsymbol{b}$ to denote vectors, where b_i represents the ith element of $\boldsymbol{b}$, and $[\boldsymbol{B}]_{i,j}$ represents the ith row and jth column of $\boldsymbol{B}$. Also, $\boldsymbol{I}_M$ is the $M \times M$ identity matrix, $\mathbf{1}_M$ is the M-length vector of ones, $\mathbf{0}_M$ is the M-length vector of zeros, $\mathrm{Diag}(\boldsymbol{b})$ is the diagonal matrix formed from the vector $\boldsymbol{b}$, $\mathrm{diag}(\boldsymbol{B})$ is the vector formed from the diagonal of matrix $\boldsymbol{B}$, $\boldsymbol{F}_N$ is the $N \times N$ unitary discrete Fourier transform (DFT) matrix, $\boldsymbol{F}_N^{1:L}$ is the matrix formed by the first L columns of $\boldsymbol{F}_N$, $\boldsymbol{f}_N^i$ is the ith column of $\boldsymbol{F}_N$, and f_N^{ij} is the $(i+1, j+1)$th element of $\boldsymbol{F}_N$. For matrices and vectors, $(\cdot)^{\mathsf{T}}$ denotes transpose, $(\cdot)^{\mathsf{H}}$ denotes conjugate transpose, $(\cdot)^*$ denotes conjugate, and $\otimes$ denotes the Kronecker product. Likewise, $\odot$, $\oslash$, and $|\cdot|^{\odot 2}$ denote element-wise multiplication, division, and absolute-value squared, respectively. Finally, the probability density function (pdf) of a multivariate complex Gaussian random vector $\boldsymbol{x}$ with mean $\widehat{\boldsymbol{x}}$ and covariance $\boldsymbol{\Sigma}$ will be denoted by $\mathcal{CN}(\boldsymbol{x}; \widehat{\boldsymbol{x}}, \boldsymbol{\Sigma})$.

II. System Model

A. Single-Carrier Block Transmission Model

We consider a single-carrier block transmission system where the transmitted frame takes the form

$$\widetilde{\boldsymbol{x}} = [\boldsymbol{x}_{\mathrm{P}}^{\mathsf{T}}, \boldsymbol{x}_{\mathrm{D}}^{\mathsf{T}}]^{\mathsf{T}}, \tag{1}$$

with $\boldsymbol{x}_{\mathrm{P}}$ a pilot frame and $\boldsymbol{x}_{\mathrm{D}}$ a data frame. For compatibility with the IEEE 802.11ad standard [2], we assume that the data frame consists of K_{D} guard-separated data blocks with guard length N_{G}, and the pilot frame consists of K_{P} pilot blocks with a cyclic-prefix (CP) structure. In particular, $\boldsymbol{x}_{\mathrm{D}} = [\boldsymbol{x}_{\mathrm{G}}^{\mathsf{T}}, \boldsymbol{x}_{\mathrm{D},1}^{\mathsf{T}}, \boldsymbol{x}_{\mathrm{G}}^{\mathsf{T}}, \ldots, \boldsymbol{x}_{\mathrm{G}}^{\mathsf{T}}, \boldsymbol{x}_{\mathrm{D},K_{\mathrm{D}}}^{\mathsf{T}}, \boldsymbol{x}_{\mathrm{G}}^{\mathsf{T}}]^{\mathsf{T}}$, where $\boldsymbol{x}_{\mathrm{G}} \in \mathbb{C}^{N_{\mathrm{G}}}$, $\boldsymbol{x}_{\mathrm{D},k} \in \mathcal{S}^{N_{\mathrm{D}}}$, and $\mathcal{S}$ is a 2^A-ary complex symbol alphabet. Note the CP structure induced by the guards. Furthermore, we assume that $\boldsymbol{x}_{\mathrm{P}} = [\boldsymbol{x}_{\mathrm{C}}, \boldsymbol{x}_{\mathrm{P},1}^{\mathsf{T}}, \ldots, \boldsymbol{x}_{\mathrm{P},K_{\mathrm{P}}}^{\mathsf{T}}]^{\mathsf{T}}$, where the last N_{C} elements of each $\boldsymbol{x}_{\mathrm{P},k} \in \mathbb{C}^M$ equal $\boldsymbol{x}_{\mathrm{C}} \in \mathbb{C}^{N_{\mathrm{C}}}$, so that the tail of each pilot block acts as the CP for the next block. Finally, we assume that $M = N_{\mathrm{D}} + N_{\mathrm{G}}$. The assumed frame structure is illustrated in Fig. 1(a).

The data sequences $\boldsymbol{x}_{D,k}$ are constructed as follows. First, $N_{\rm b}$ information bits $\boldsymbol{b} \triangleq [b_1, \ldots, b_{N_{\rm b}}]^{\mathsf{T}}$ are coded and then interleaved, yielding the coded bits $\boldsymbol{c} \in \{0,1\}^{AK_{\rm D}N_{\rm D}}$ and a code rate of $R = \frac{N_{\rm b}}{AK_{\rm D}N_{\rm D}}$. Next, the coded bits are partitioned into $K_{\rm D}N_{\rm D}$ groups of A bits, $\boldsymbol{c} \triangleq [\boldsymbol{c}_0^{\mathsf{T}}, \ldots, \boldsymbol{c}_{K_{\rm D}N_{\rm D}-1}^{\mathsf{T}}]^{\mathsf{T}}$, where each group $\boldsymbol{c}_n \triangleq [c_{n,1}, \ldots, c_{n,A}]^{\mathsf{T}}$ determines the value of one data symbol. By partitioning the $K_{\rm D}N_{\rm D}$ data symbols into $K_{\rm D}$ blocks of $N_{\rm D}$ symbols, one obtains the data sequences $\boldsymbol{x}_{{\rm D},k}$ for $k = 1, \ldots, K_{\rm D}$.

B. Propagation and Few-Bit ADC Model

The frame $\widetilde{\boldsymbol{x}}$ is modulated using a square-root raised-cosine pulse, upconverted, propagated through a noisy and frequency-selective channel (using possibly many antennas with analog beamforming at the transmitter and/or receiver), downconverted, filtered with a square-root raised cosine pulse, and sampled at the baud rate. We will assume that the beamformed baseband channel impulse response, $\boldsymbol{h} \triangleq [h_0, \ldots, h_{L-1}]^{\mathsf{T}}$, has length $L \leq \min\{N_{\rm C}, N_{\rm G}\} - 1$ and is invariant during the transmission of $\widetilde{\boldsymbol{x}}$. In this case, after discarding the received samples corresponding to the first $\boldsymbol{x}_{\rm C}$ and $\boldsymbol{x}_{\rm G}$ sequences, the unquantized received samples can be collected into the matrix

$$\boldsymbol{U} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{W}, \tag{2}$$

where $K \triangleq K_{\rm P} + K_{\rm D}$. In (2), $\boldsymbol{H} \in \mathbb{C}^{M \times M}$ is the circulant matrix with first column $[\boldsymbol{h}^{\mathsf{T}}\, \boldsymbol{0}_{M-L}^{\mathsf{T}}]^{\mathsf{T}}$, $\boldsymbol{W} \in \mathbb{C}^{M \times K}$ contains additive white Gaussian noise (AWGN) with variance σ_w^2, which is assumed to be known,[1] and the kth column of $\boldsymbol{X} \in \mathbb{C}^{M \times K}$ equals $\boldsymbol{x}_{{\rm P},k}$ when $k \in \{1, \ldots, K_{\rm P}\}$ or $[\boldsymbol{x}_{{\rm D},k-K_{\rm P}}^{\mathsf{T}}, \boldsymbol{x}_{\rm G}^{\mathsf{T}}]^{\mathsf{T}}$ when $k > K_{\rm P}$. Likewise, we can write (2) in vectorized form as

$$\boldsymbol{u} = (\boldsymbol{I}_K \otimes \boldsymbol{H})\boldsymbol{x} + \boldsymbol{w}, \tag{3}$$

with $\boldsymbol{u} \triangleq \mathrm{vec}(\boldsymbol{U})$, $\boldsymbol{x} \triangleq \mathrm{vec}(\boldsymbol{X})$, $\boldsymbol{w} \triangleq \mathrm{vec}(\boldsymbol{W})$, and $\otimes$ denoting the Kronecker product. It can be shown that $\boldsymbol{x}$ equals $\widetilde{\boldsymbol{x}}$ with the first $\boldsymbol{x}_{\rm C}$ and $\boldsymbol{x}_{\rm G}$ sequences removed.

The output of the few-bit ADC is modeled as

$$\boldsymbol{y} = \mathcal{Q}\big(\boldsymbol{u}\big), \tag{4}$$

where the quantization $\mathcal{Q}(\cdot)$ applies component-wise. Although not required by our methodology, we will assume in our numerical experiments that b-bit uniform mid-rise quantization [46] is separately applied to the real and imaginary parts, i.e.,

$$\begin{aligned} y_m = {} & \mathrm{sign}(\mathrm{Re}(u_m)) \left(\min\left\{ \left\lceil \frac{|\mathrm{Re}(u_m)|}{\triangle_{\rm Re}} \right\rceil, 2^{b-1} \right\} - \frac{1}{2} \right) \\ & + \mathrm{j}\, \mathrm{sign}(\mathrm{Im}(u_m)) \left(\min\left\{ \left\lceil \frac{|\mathrm{Im}(u_m)|}{\triangle_{\rm Im}} \right\rceil, 2^{b-1} \right\} - \frac{1}{2} \right), \end{aligned} \tag{5}$$

where $\triangle_{\rm Re} \triangleq \sqrt{\mathbb{E}\big[\mathrm{Re}(u_m)^2\big]}\triangle_b$, $\triangle_{\rm Im} \triangleq \sqrt{\mathbb{E}\big[\mathrm{Im}(u_m)^2\big]}\triangle_b$, and $\triangle_{\rm b}$ is chosen to minimize the mean-squared error (MSE) $\mathbb{E}\big[|y_m - u_m|^2\big]$ under Gaussian u_m. The average powers $\mathbb{E}\big[\mathrm{Re}(u_m)^2\big]$ and $\mathbb{E}\big[\mathrm{Im}(u_m)^2\big]$ can be measured by analog circuits before the ADC. When $b > 1$, such measurements are typically performed as part of automatic gain control.

[1]The noise variance could be estimated using the EM-PBiGAMP procedure described in [37], but we leave the verification of this approach to future work. See [45] for AWGN-variance learning under 1-bit quantization, referred to as the "probit link" in the context of binary classification.

C. Channel Model for Propagation

For signal propagation, we used the 60 GHz wireless local area network (WLAN) channel model adopted by the IEEE 802.11ad task group [44], which was a result of extensive channel measurement studies in [28]. It specifies that the continuous-space/time channel impulse response $h(t; \phi_{\rm tx}, \theta_{\rm tx}, \phi_{\rm rx}, \theta_{\rm rx})$, as a function of the lag t, the azimuth angles $(\phi_{\rm tx}, \phi_{\rm rx})$, and the elevation angles $(\theta_{\rm tx}, \theta_{\rm rx})$, takes the form

$$\begin{aligned} & h(t; \phi_{\rm tx}, \theta_{\rm tx}, \phi_{\rm rx}, \theta_{\rm rx}) \\ & = \sum_{i=1}^{I} \alpha^{(i)} C^{(i)}\Big(t - \tau^{(i)}; \phi_{\rm tx} - \Phi_{\rm tx}^{(i)}, \theta_{\rm tx} - \Theta_{\rm tx}^{(i)}, \\ & \quad \phi_{\rm rx} - \Phi_{\rm rx}^{(i)}, \theta_{\rm rx} - \Theta_{\rm rx}^{(i)}\Big) \end{aligned} \tag{6a}$$

$$\begin{aligned} & C^{(i)}(t; \phi_{\rm tx}, \theta_{\rm tx}, \phi_{\rm rx}, \theta_{\rm rx}) \\ & = \sum_{u=1}^{U^{(i)}} \alpha^{(i,u)} \delta\big(t - \tau^{(i,u)}\big) \delta\big(\phi_{\rm tx} - \Phi_{\rm tx}^{(i,u)}\big) \delta\big(\theta_{\rm tx} - \Theta_{\rm tx}^{(i,u)}\big) \\ & \quad \times \delta\big(\phi_{\rm rx} - \Phi_{\rm rx}^{(i,u)}\big) \delta\big(\theta_{\rm rx} - \Theta_{\rm rx}^{(i,u)}\big), \end{aligned} \tag{6b}$$

where

- $\alpha^{(i)}$ and $C^{(i)}(t; \phi_{\rm tx}, \theta_{\rm tx}, \phi_{\rm rx}, \theta_{\rm rx})$ are the gain and channel impulse response of the ith cluster, respectively,
- $\tau^{(i)}$, $\Phi_{\rm tx}^{(i)}$, $\Theta_{\rm tx}^{(i)}$, $\Phi_{\rm rx}^{(i)}$, $\Theta_{\rm rx}^{(i)}$ are the delay-angle coordinates of the ith cluster,
- $\alpha^{(i,u)}$ is the gain of the uth ray of the ith cluster,
- $\tau^{(i,u)}$, $\Phi_{\rm tx}^{(i,u)}$, $\Theta_{\rm tx}^{(i,u)}$, $\Phi_{\rm rx}^{(i,u)}$, $\Theta_{\rm rx}^{(i,u)}$ are the relative delay-angle coordinates of the uth ray of the ith cluster,
- I is the number of clusters and $U^{(i)}$ is the number of rays in the ith cluster, and
- $\delta(\cdot)$ is the Dirac delta.

The discrete-time impulse response coefficients $\{h_l\}$ are constructed from $h(t; \phi_{\rm tx}, \theta_{\rm tx}, \phi_{\rm rx}, \theta_{\rm rx})$ via pulse-shaping and beamforming, i.e.,

$$\begin{aligned} h_l = {} & \int h(t; \phi_{\rm tx}, \theta_{\rm tx}, \phi_{\rm rx}, \theta_{\rm rx}) g(lT - t) \\ & \times b_{\rm tx}(\phi_{\rm tx}, \theta_{\rm tx}) b_{\rm rx}(\phi_{\rm rx}, \theta_{\rm rx})\, \mathrm{d}t\, \mathrm{d}\phi_{\rm tx}\, \mathrm{d}\theta_{\rm tx}\, \mathrm{d}\phi_{\rm rx}\, \mathrm{d}\theta_{\rm rx}, \end{aligned} \tag{7}$$

where $g(\cdot)$ is the pulse shape specified in the 802.11ad standard (i.e., raised-cosine with rolloff 0.25), T is the baud interval, and $b_{\rm tx}(\phi_{\rm tx}, \theta_{\rm tx})$ and $b_{\rm rx}(\phi_{\rm rx}, \theta_{\rm rx})$ are beam responses.

Based on extensive physical channel measurements, statistical models for the 60 GHz WLAN channel parameters were proposed in [44], and Matlab code to generate realizations from this model (including optimized analog beamforming) was provided in [47]. Typical realizations of the resulting $\{|h_l|\}_{l=0}^{L-1}$ from the "conference room" environment are shown in Figs. 2(a)–(b), which show that the channel taps are approximately sparse. The channel power-delay profile (PDP), $\mathbb{E}\{|h_l|^2\}$ versus l, is plotted

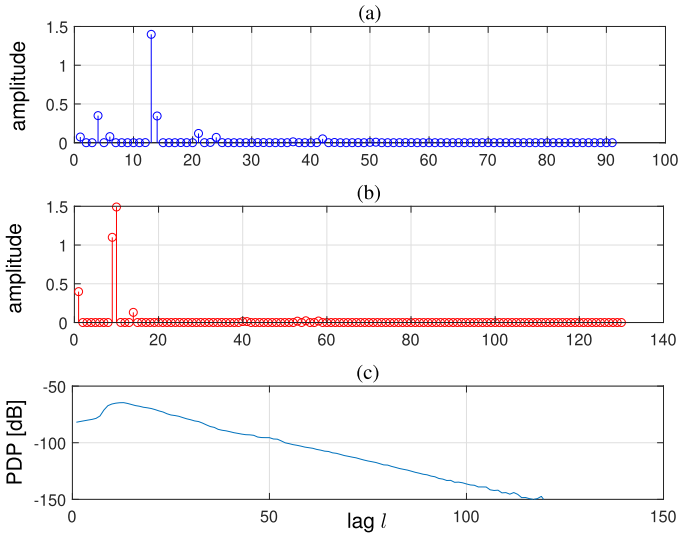


Fig. 2. For the 802.11ad 60 GHz "conference room" channel, typical realizations of $|h_l|$ versus l are shown in (a) and (b), and the power-delay profile is shown in (c).

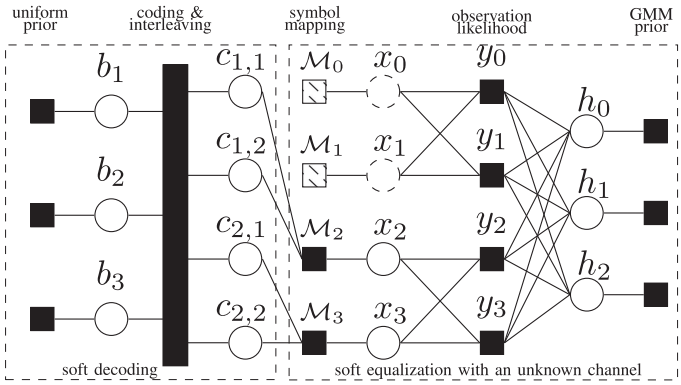


Fig. 3. The factor graph corresponding to a toy example with $N_{\mathrm{b}} = 3$ information bits $\{b_i\}$, 4 interleaved/coded bits $\{c_{n,a}\}$, $A = 2$ bits/symbol, $N_{\mathrm{D}} = 2$ data symbols per block, $N_{\mathrm{G}} = 0$ guard symbols per block, $K_{\mathrm{P}} = 1$ pilot blocks, $K_{\mathrm{D}} = 1$ data blocks, block length $M = N_{\mathrm{D}} + N_{\mathrm{G}} = 2$, pilot symbols x_0 and x_1, data symbols x_2 and x_3, and $L = 3$ channel taps. The node y_m represents $p(y_m|z_m)$ and the node $\mathcal{M}_n$ represents the bit-to-symbol mapping for data symbols or the indicator pmf for pilot symbols.

in Fig. 2(c), with the expectation approximated by an average of 50 000 realizations. There it can be seen that the PDP decays exponentially with lag l, i.e., the index into $\boldsymbol{h}$.

D. Channel Model for Estimation

The channel model as given in (7) is difficult to directly exploit for channel estimation. Therefore, for channel estimation, we propose to use a D-state Gaussian-mixture model (GMM) for the channel vector $\boldsymbol{h}$, as suggested in [39] for $D = 2$. For general $D \geq 1$, the GMM specifies a pdf of the form

$$p(\boldsymbol{h};\boldsymbol{\lambda},\boldsymbol{\nu}) = \prod_{l=0}^{L-1} p(h_l;\boldsymbol{\lambda}_l,\boldsymbol{\nu}_l) \tag{8a}$$

$$p(h_l;\boldsymbol{\lambda}_l,\boldsymbol{\nu}_l) = \sum_{d=1}^{D} \lambda_{l,d}\, \mathcal{CN}(h_l;0,\nu_{l,d}), \tag{8b}$$

where $\lambda_{l,d} \geq 0$ and $\nu_{l,d} > 0$ are the weight and variance of the dth mixture component of the l tap, and $\sum_{d=1}^{D} \lambda_{l,d} = 1\,\forall l$. Also, $\boldsymbol{\lambda}_l \triangleq [\lambda_{l,1},\ldots,\lambda_{l,D}]^{\mathsf{T}}$ and $\boldsymbol{\lambda} \triangleq [\boldsymbol{\lambda}_0^{\mathsf{T}},\ldots,\boldsymbol{\lambda}_{L-1}^{\mathsf{T}}]^{\mathsf{T}}$, with similar definitions for $\boldsymbol{\nu}_l$ and $\boldsymbol{\nu}$. In principle, the GMM parameters, $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$, could be empirically estimated from a corpus of training data using the standard EM-based approach to fitting a GMM [48, p. 435]. As an alternative, these parameters can be estimated online from the quantized measurements $\boldsymbol{y}$ using the EM-AMP-based method described in Section III-E.

III. Turbo Equalization With PBiGAMP

Our principle goal is to infer the information bits $\boldsymbol{b}$ from the few-bit measurements $\boldsymbol{y}$ under the block-transmission model from Section II-A, the few-bit ADC model from Section II-B, and the GMM channel model from Section II-D. In particular, we aim to compute the marginal posterior probabilities $\{p(b_i|\boldsymbol{y})\}_{i=1}^{N_{\mathrm{b}}}$, which can be decomposed as

$$p(b_i|\boldsymbol{y}) = \sum_{\boldsymbol{b}_{-i}} p(\boldsymbol{b}|\boldsymbol{y}) = \sum_{\boldsymbol{b}_{-i}} \frac{p(\boldsymbol{y}|\boldsymbol{b})p(\boldsymbol{b})}{p(\boldsymbol{y})} \propto \sum_{\boldsymbol{b}_{-i}} p(\boldsymbol{y}|\boldsymbol{b}) \tag{9}$$

$$= \sum_{\boldsymbol{b}_{-i},\boldsymbol{x},\boldsymbol{c}} \int_{\mathbb{C}^L} p(\boldsymbol{y}|\boldsymbol{h},\boldsymbol{x})p(\boldsymbol{h})p(\boldsymbol{x}|\boldsymbol{c})p(\boldsymbol{c}|\boldsymbol{b})\,\mathrm{d}\boldsymbol{h} \tag{10}$$

$$= \sum_{\boldsymbol{b}_{-i},\boldsymbol{c}} p(\boldsymbol{c}|\boldsymbol{b}) \sum_{\boldsymbol{x}} \int_{\mathbb{C}^L} \left[\prod_{m=1}^{MK} p(y_m|\boldsymbol{h},\boldsymbol{x})\right] \left[\prod_{l=0}^{L-1} p(h_l)\right] \mathrm{d}\boldsymbol{h} \times \left[\prod_{k=1}^{K_{\mathrm{D}}} \prod_{n=0}^{N_{\mathrm{D}}-1} p(x_{(K_{\mathrm{P}}+k-1)M+n}|\boldsymbol{c}_{(k-1)N_{\mathrm{D}}+n})\right], \tag{11}$$

for $\boldsymbol{b}_{-i} \triangleq [b_1,\ldots,b_{i-1},b_{i+1},\ldots,b_{N_{\mathrm{b}}}]^{\mathsf{T}}$. Above, (9) is due to Bayes rule and the assumption that the information bits $\boldsymbol{b}$ are uniformly distributed; (10) is due to the dependency relationships among the random vectors $\boldsymbol{y}$, $\boldsymbol{h}$, $\boldsymbol{x}$, $\boldsymbol{c}$, and $\boldsymbol{b}$; and (11) is due to the separable nature of $p(\boldsymbol{y}|\boldsymbol{h},\boldsymbol{x})$, $p(\boldsymbol{h})$, and $p(\boldsymbol{x}|\boldsymbol{c})$. In particular, the pmfs $p(x_{(K_{\mathrm{P}}+k-1)M+n}|\boldsymbol{c}_{(k-1)N_{\mathrm{D}}+n})$ for $k = 1,\ldots,K_{\mathrm{D}}$ and $n = 0,\ldots,N_{\mathrm{D}}-1$ are determined by the bit-to-symbol mapping, and the likelihood function $p(y_m|\boldsymbol{h},\boldsymbol{x})$ can be obtained from (3)–(4). Details are provided in the sequel.

The structure in (11) can be visualized using the bipartite factor graph shown in Fig. 3, where the solid rectangles represent the pdf factors and the open circles represent the variable nodes. We find it convenient to partition the factor graph into two subgraphs: the left subgraph corresponds to soft decoding and the right subgraph corresponds to soft equalization with an unknown channel.

A. Belief Propagation

The posterior bit marginals $\{p(b_i|\boldsymbol{y})\}_{i=1}^{N_{\mathrm{b}}}$ can in principle be computed from (11), but doing so is impractical from the standpoint of complexity. A practical alternative is to perform

belief-propagation (BP) using the sum-product algorithm (SPA) [49], which passes messages along the edges of the factor graph in Fig. 3. For discrete-valued variables like $b_i, c_{n,a}, x_n$, these messages come in the form of pmfs, while for continuous variables like h_l, these messages come in the form of pdfs. When there are no loops (i.e., cycles) in the factor graph, BP computes exact marginals. But Fig. 3 has loops, and so BP computes only approximate marginals. This is to be expected, given that exact inference in loopy graphs is NP hard [50]. Still, loopy BP often gives very good results, and so it has become popular for, e.g., turbo decoding, LDPC decoding, turbo equalization, inference of Markov random fields, multiuser detection, and compressive sensing.

Exact implementation of the SPA is intractable for the soft-equalization subgraph in Fig. 3. For exact SPA, the messages in and out of the h_l nodes would take the form of Gaussian mixtures, with a mixture order that grows exponentially in the iterations. As an alternative, one might consider passing only Gaussian approximations of these problematic SPA messages, an approach known as expectation propagation (EP) [51]. But since there are MKL edges between the $\{h_l\}$ and $\{y_m\}$ nodes in Fig. 3, the per-symbol complexity of EP would be $O(L)$, which contrasts with the $O(\log L)$ complexity of FFT processing. Also, the fixed-points of EP are generally not well understood.

B. Background on PBiGAMP

We now briefly provide some background on PBiGAMP, since many readers may not be familiar with the algorithm. PBiGAMP [37] is a computationally efficient approach to approximating the marginal posteriors of independent random variables $\{\mathsf{x}_n\}_{n=0}^{N-1}$ and $\{\mathsf{h}_l\}_{l=0}^{L-1}$ from measurements $\boldsymbol{y} = [y_0, \ldots, y_{P-1}]^{\mathsf{T}}$ generated under a likelihood of the form

$$p_{\mathsf{y}|\mathsf{z}}(\boldsymbol{y}|\boldsymbol{z}) = \prod_{m=0}^{P-1} p_{\mathsf{y}_m|\mathsf{z}_m}(y_m|z_m) \tag{12a}$$

$$\mathsf{z}_m = \sum_{n=0}^{N-1}\sum_{l=0}^{L-1} \mathsf{x}_n z_m^{(n,l)} \mathsf{h}_l, \tag{12b}$$

where $z_m^{(n,l)}$ are known parameters. Throughout this subsection, we typeset random variables in san-serif font (e.g., y_m) and non-random variables in serif font (e.g., y_m) for clarity. Note that, in (12), z_m can be interpreted as noiseless bilinear measurements of the random vectors $\mathsf{x} \triangleq [\mathsf{x}_0, \ldots, \mathsf{x}_{N-1}]^{\mathsf{T}}$ and $\mathsf{h} \triangleq [\mathsf{h}_0, \ldots, \mathsf{h}_{L-1}]^{\mathsf{T}}$, and $p_{\mathsf{y}_m|\mathsf{z}_m}(y_m|z_m)$ can be interpreted as a noisy measurement channel. Applications of (12) include matrix compressive sensing, self-calibration, blind deconvolution, and joint channel/symbol estimation.

The PBiGAMP algorithm from [37] is summarized in Table I. There, the priors on x_n and h_l are denoted by $p_{\mathsf{x}_n}(x_n)$ and $p_{\mathsf{h}_l}(h_l)$, respectively. The approximate marginal posteriors, denoted by $p_{\mathsf{x}_n|\mathsf{q}_n}(x_n|\widehat{q}_n;\nu_n^{\mathsf{q}})$ and $p_{\mathsf{h}_l|\mathsf{r}_l}(h_l|\widehat{r}_l;\nu_l^{\mathsf{r}})$, are specified in lines (D2)–(D3). Here, $\widehat{q}_n, \nu_n^{\mathsf{q}}, \widehat{r}_l, \nu_l^{\mathsf{r}}$ are quantities computed iteratively by PBiGAMP.

In [37], PBiGAMP was derived as a computationally efficient approximation of the SPA for the likelihood model (12), assuming that $z_m^{(n,l)}$ are independent realizations of a zero-mean Gaussian random variable. This approximation is, in fact, exact in the large-system limit (i.e., $P, N, L \to \infty$ with fixed N/P and L/P). In [52], PBiGAMP was analyzed using the replica method from statistical physics. There it was shown that the large-system-limit performance of PBiGAMP can be accurately predicted by a scalar state-evolution. For the case of i.i.d. Bernoulli-Gaussian x_n and h_l, this state evolution was studied in detail and found to exhibit a sharp "phase-transition" behavior. Moreover, for certain combinations of measurement rates (i.e., N/P and L/P) and sparsity rates on x_n and h_l, PBiGAMP was shown to converge to the MMSE estimates of $\boldsymbol{x}$ and $\boldsymbol{h}$. For other, more difficult, combinations of measurement and sparsity rates, PBiGAMP may not yield accurate estimates. However, it is conjectured that no other polynomial-time method will yield accurate estimates in that case [52].

TABLE I
THE SCALAR-VARIANCE PBiGAMP ALGORITHM FROM [37]

Definitions:

$$p_{\mathsf{z}_m|\mathsf{p}_m}(z\,|\,\widehat{p};\nu^{\mathsf{p}}) \triangleq \frac{p_{\mathsf{y}_m|\mathsf{z}_m}(y_m\,|\,z)\,\mathcal{CN}(z;\widehat{p},\nu^{\mathsf{p}})}{\int p_{\mathsf{y}_m|\mathsf{z}_m}(y_m\,|\,z')\,\mathcal{CN}(z';\widehat{p},\nu^{\mathsf{p}})\,\mathrm{d}z'} \tag{D1}$$

$$p_{\mathsf{h}_l|\mathsf{r}_l}(h\,|\,\widehat{r};\nu^{\mathsf{r}}) \triangleq \frac{p_{\mathsf{h}_l}(h)\,\mathcal{CN}(\widehat{r};h,\nu^{\mathsf{r}})}{\int p_{\mathsf{h}_l}(h')\,\mathcal{CN}(\widehat{r};h',\nu^{\mathsf{r}})\,\mathrm{d}h'} \tag{D2}$$

$$p_{\mathsf{x}_n|\mathsf{q}_n}(x\,|\,\widehat{q};\nu^{\mathsf{q}}) \triangleq \frac{p_{\mathsf{x}_n}(x)\,\mathcal{CN}(\widehat{q};x,\nu^{\mathsf{q}})}{\int p_{\mathsf{x}_n}(x')\,\mathcal{CN}(\widehat{q};x',\nu^{\mathsf{q}})\,\mathrm{d}x'} \tag{D3}$$

Initialization:

$$\forall m: \widehat{s}_m[0] = 0 \tag{I1}$$

$$\forall n,l: \text{choose} \quad \widehat{x}_n[1], \nu^{\mathsf{x}}[1], \widehat{h}_l[1], \nu^{\mathsf{h}}[1] \tag{I2}$$

For $t = 1, \ldots T_{\max}$

$$\forall n: \widehat{\boldsymbol{z}}^{(n,*)}[t] = \textstyle\sum_{l=0}^{L-1} \boldsymbol{z}^{(n,l)} \widehat{h}_l[t] \tag{R1}$$

$$\forall l: \widehat{\boldsymbol{z}}^{(*,l)}[t] = \textstyle\sum_{n=0}^{N-1} \widehat{x}_n[t] \boldsymbol{z}^{(n,l)} \tag{R2}$$

$$\widehat{\boldsymbol{z}}^{(*,*)}[t] = \textstyle\sum_{n=0}^{N-1} \widehat{x}_n[t] \widehat{\boldsymbol{z}}^{(n,*)}[t] \text{ or } \sum_{l=0}^{L-1} \widehat{h}_l[t] \widehat{\boldsymbol{z}}^{(*,l)}[t] \tag{R3}$$

$$\overline{\nu}^{\mathsf{p}}[t] = \tfrac{1}{P}\big(\nu^{\mathsf{x}}[t] \textstyle\sum_{n=0}^{N-1} \|\widehat{\boldsymbol{z}}^{(n,*)}[t]\|^2 + \nu^{\mathsf{h}}[t] \sum_{l=0}^{L-1} \|\widehat{\boldsymbol{z}}^{(*,l)}[t]\|^2\big) \tag{R4}$$

$$\nu^{\mathsf{p}}[t] = \overline{\nu}^{\mathsf{p}}[t] + \nu^{\mathsf{x}}[t]\nu^{\mathsf{h}}[t] \tfrac{1}{P} \textstyle\sum_{n=0}^{N-1} \sum_{l=0}^{L-1} \|\boldsymbol{z}^{(n,l)}[t]\|^2 \tag{R5}$$

$$\widehat{\boldsymbol{p}}[t] = \widehat{\boldsymbol{z}}^{(*,*)}[t] - \widehat{\boldsymbol{s}}[t-1]\overline{\nu}^{\mathsf{p}}[t] \tag{R6}$$

$$\nu^{z}[t] = \tfrac{1}{P} \textstyle\sum_{m=0}^{P-1} \mathrm{var}\{\mathsf{z}_m \,|\, \mathsf{p}_m = \widehat{p}_m[t]; \nu^{\mathsf{p}}[t]\} \tag{R7}$$

$$\forall m: \widehat{z}_m[t] = \mathbb{E}[\mathsf{z}_m \,|\, \mathsf{p}_m = \widehat{p}_m[t]; \nu^{\mathsf{p}}[t]] \tag{R8}$$

$$\nu^{\mathsf{s}}[t] = (1 - \nu^{z}[t]/\nu^{\mathsf{p}}[t])/\nu^{\mathsf{p}}[t] \tag{R9}$$

$$\widehat{\boldsymbol{s}}[t] = (\widehat{\boldsymbol{z}}[t] - \widehat{\boldsymbol{p}}[t])/\nu^{\mathsf{p}}[t] \tag{R10}$$

$$\nu^{\mathsf{r}}[t] = \big(\nu^{\mathsf{s}}[t] \tfrac{1}{L} \textstyle\sum_{l=0}^{L-1} \|\widehat{\boldsymbol{z}}^{(*,l)}[t]\|^2\big)^{-1} \tag{R11}$$

$$\forall l: \widehat{r}_l[t] = \widehat{h}_l[t] + \nu^{\mathsf{r}}[t] \widehat{\boldsymbol{z}}^{(*,l)\mathsf{H}}[t] \widehat{\boldsymbol{s}}[t] - \nu^{\mathsf{r}}[t]\nu^{\mathsf{s}}[t]\nu^{\mathsf{x}}[t]\widehat{h}_l[t] \textstyle\sum_{n=0}^{N-1} \|\boldsymbol{z}^{(n,l)}\|^2 \tag{R12}$$

$$\nu^{\mathsf{q}}[t] = \big(\nu^{\mathsf{s}}[t] \tfrac{1}{N} \textstyle\sum_{n=0}^{N-1} \|\widehat{\boldsymbol{z}}^{(n,*)}[t]\|^2\big)^{-1} \tag{R13}$$

$$\forall n: \widehat{q}_n[t] = \widehat{x}_n[t] + \nu^{\mathsf{q}}[t] \widehat{\boldsymbol{z}}^{(n,*)\mathsf{H}}[t] \widehat{\boldsymbol{s}}[t] - \nu^{\mathsf{q}}[t]\nu^{\mathsf{s}}[t]\nu^{\mathsf{h}}[t]\widehat{x}_n[t] \textstyle\sum_{l=0}^{L-1} \|\boldsymbol{z}^{(n,l)}\|^2 \tag{R14}$$

$$\nu^{\mathsf{h}}[t+1] = \tfrac{1}{L} \textstyle\sum_{l=0}^{L-1} \mathrm{var}\{\mathsf{h}_l \,|\, \mathsf{r}_l = \widehat{r}_l[t]; \nu^{\mathsf{r}}[t]\} \tag{R15}$$

$$\forall l: \widehat{h}_l[t+1] = \mathbb{E}[\mathsf{h}_l \,|\, \mathsf{r}_l = \widehat{r}_l[t]; \nu^{\mathsf{r}}[t]] \tag{R16}$$

$$\nu^{\mathsf{x}}[t+1] = \tfrac{1}{N} \textstyle\sum_{n=0}^{N-1} \mathrm{var}\{\mathsf{x}_n \,|\, \mathsf{q}_n = \widehat{q}_n[t]; \nu^{\mathsf{q}}[t]\} \tag{R17}$$

$$\forall n: \widehat{x}_n[t+1] = \mathbb{E}[\mathsf{x}_n \,|\, \mathsf{q}_n = \widehat{q}_n[t]; \nu^{\mathsf{q}}[t]] \tag{R18}$$

end

C. Soft Equalization via PBiGAMP

In this section, we describe how PBiGAMP can be applied to soft equalization of SC block transmissions over unknown FS channels measured by few-bit ADCs.

We begin by adapting the PBiGAMP likelihood model (12) to the few-bit SC block-transmission model (3)–(4). First, we

write the circulant channel matrix as $\boldsymbol{H} = \sum_{l=0}^{L-1} h_l \boldsymbol{J}_l$, where $\boldsymbol{J}_l \in \mathbb{R}^{M\times M}$ is the l-circulant delay matrix. Then (4) becomes

$$y_m = \mathcal{Q}\left(\sum_{l=0}^{L-1}\sum_{n=0}^{MK-1} h_l[\boldsymbol{I}_K \otimes \boldsymbol{J}_l]_{m,n} x_n + w_m\right), \quad (13)$$

where $[\cdot]_{m,n}$ extracts the mth row and nth column of its matrix argument. From (12) and (13), we can readily identify the PBiGAMP quantities

$$z_m^{(n,l)} = [\boldsymbol{I}_K \otimes \boldsymbol{J}_l]_{m,n} \quad (14)$$

$$p_{\mathsf{y}_m|\mathsf{z}_m}(y_m|z_m) \triangleq \Pr\{y_m = \mathcal{Q}(z_m + \mathsf{w}_m)\} \quad (15)$$

$$= \int_{\mathcal{Q}^{-1}(y_m)} \mathcal{CN}\big(w; z_m, \sigma_w^2\big)\,\mathrm{d}w, \quad (16)$$

where $\mathcal{Q}^{-1}(y_m) \subset \mathbb{C}$ is the region quantized to y_m. We also identify the PBiGAMP dimensions $P = N = MK$.

For PBiGAMP's prior on h_l, we assign the GMM from (8). For PBiGAMP's prior on x_n, we treat the indices n of data symbols differently from those of pilot and guard symbols. For the data indices $n \in \{(K_\mathrm{P}+k-1)M, \ldots, (K_\mathrm{P}+k-1)M + N_\mathrm{D} - 1\}_{k=1}^{K_\mathrm{D}}$, we assign

$$p_{\mathsf{x}_n}(x_n) = \sum_{j=1}^{2^A} \gamma_{n,j}\,\delta(x_n - s^{(j)}), \quad (17)$$

where $\delta(\cdot)$ is the Dirac delta, $\{s^{(1)}, \ldots, s^{(2^A)}\} \triangleq \mathcal{S}$ is the data-symbol alphabet, and $\gamma_{n,j} = \Pr\{\mathsf{x}_n = s^{(j)}\}$ is the prior data-symbol pmf, which depends on the decoder outputs as described below. For pilot indices $n = 0, \ldots, K_\mathrm{P}M - 1$ and guard indices $n \in \{(K_\mathrm{P}+k-1)M + N_\mathrm{D}, \ldots, (K_\mathrm{P}+k)M - 1\}_{k=1}^{K_\mathrm{D}}$, we assign the trivial prior $p_{\mathsf{x}_n}(x) = \delta(x - x_n)$ because the pilots and guards take on known deterministic values. Note that, although the data symbols x_n are discrete, PBiGAMP treats them as continuous random variables in $\mathbb{C}$.

The data-symbol pmf $\{\gamma_{n,j}\}_{j=1}^{2^A}$ is determined by the coded-bit priors $\Pr\{\mathsf{c}_{n,a} = c_a^{(j)}\}$ coming from the soft decoder, i.e.,

$$\gamma_{n,j} \triangleq \Pr\{\mathsf{x}_n = s^{(j)}\} = \sum_{j'=1}^{2^A} \Pr\{\mathsf{x}_n = s^{(j)}, \mathsf{c}_n = \boldsymbol{c}^{(j')}\} \quad (18)$$

$$= \sum_{j'=1}^{2^A} \underbrace{\Pr\{\mathsf{x}_n = s^{(j)} | \mathsf{c}_n = \boldsymbol{c}^{(j')}\}}_{\delta_{j-j'}} \Pr\{\mathsf{c}_n = \boldsymbol{c}^{(j')}\} \quad (19)$$

$$= \Pr\{\mathsf{c}_n = \boldsymbol{c}^{(j)}\} = \prod_{a=1}^{A} \Pr\{\mathsf{c}_{n,a} = c_a^{(j)}\}, \quad (20)$$

where $\boldsymbol{c}^{(j)} = [c_1^{(j)}, \ldots, c_A^{(j)}]^\mathsf{T} \in \{0,1\}^A$ is the coded-bit sequence corresponding to the symbol value $s^{(j)}$, and δ_j is the Kronecker delta sequence.

We are now ready to apply PBiGAMP from Table I. In the sequel, we omit the iteration index "$[t]$" for brevity. From (14) and $\boldsymbol{z}^{(n,l)} \triangleq [z_0^{(n,l)}, \ldots, z_{MK-1}^{(n,l)}]^\mathsf{T}$, lines (R1)–(R3) of Table I become

$$\widehat{\boldsymbol{z}}^{(n,*)} = \sum_{l=0}^{L-1} \widehat{h}_l [\boldsymbol{I}_K \otimes \boldsymbol{J}_l]_{:,n} = [\boldsymbol{I}_K \otimes \widehat{\boldsymbol{H}}]_{:,n} \quad (21)$$

$$\widehat{\boldsymbol{z}}^{(*,l)} = \sum_{n=0}^{MK-1} \widehat{x}_n [\boldsymbol{I}_K \otimes \boldsymbol{J}_l]_{:,n} = \mathrm{vec}\big(\boldsymbol{J}_l \widehat{\boldsymbol{X}}\big) \quad (22)$$

$$\widehat{\boldsymbol{z}}^{(*,*)} = \sum_{l=0}^{L-1} \widehat{h}_l\, \mathrm{vec}\big(\boldsymbol{J}_l \widehat{\boldsymbol{X}}\big) = \mathrm{vec}\big(\widehat{\boldsymbol{H}}\widehat{\boldsymbol{X}}\big), \quad (23)$$

where $[\cdot]_{:,n}$ extracts the nth column of its matrix argument, $\widehat{\boldsymbol{H}} = \sum_{l=0}^{L-1} \widehat{h}_l \boldsymbol{J}_l \in \mathbb{C}^{M\times M}$ is the circulant matrix with first column $[\widehat{\boldsymbol{h}}^\mathsf{T}\ \boldsymbol{0}_{M-L}^\mathsf{T}]^\mathsf{T}$, and $\widehat{\boldsymbol{X}} \in \mathbb{C}^{M\times K}$ is such that $\widehat{\boldsymbol{x}} = \mathrm{vec}(\widehat{\boldsymbol{X}})$. Given (21)–(23), the structure of $\widehat{\boldsymbol{H}}$ and $\boldsymbol{J}_l$ imply

$$\|\widehat{\boldsymbol{z}}^{(n,*)}\|^2 = \|\widehat{\boldsymbol{h}}\|^2\ \forall n \quad (24)$$

$$\|\widehat{\boldsymbol{z}}^{(*,l)}\|^2 = \|\widehat{\boldsymbol{x}}\|^2 = \|\widehat{\boldsymbol{X}}\|_F^2\ \forall l \quad (25)$$

$$\|\boldsymbol{z}^{(n,l)}\|^2 = 1\ \forall n, l. \quad (26)$$

With (23)–(26), PBiGAMP steps (R4)–(R6) reduce to

$$\overline{\nu}^\mathsf{p} = \nu^\mathsf{x}\|\widehat{\boldsymbol{h}}\|^2 + \frac{L}{MK}\nu^\mathsf{h}\|\widehat{\boldsymbol{x}}\|^2 \quad (27)$$

$$\nu^\mathsf{p} = \overline{\nu}^\mathsf{p} + L\nu^\mathsf{x}\nu^\mathsf{h} \quad (28)$$

$$\widehat{\boldsymbol{p}} = \mathrm{vec}(\widehat{\boldsymbol{H}}\widehat{\boldsymbol{X}}) - \overline{\nu}^\mathsf{p}\widehat{\boldsymbol{s}}. \quad (29)$$

Furthermore, because $\widehat{\boldsymbol{H}}$ is circulant, its eigendecomposition takes the form

$$\widehat{\boldsymbol{H}} = \sqrt{M}\boldsymbol{F}_M^\mathsf{H}\,\mathrm{Diag}(\boldsymbol{F}_M^{1:L}\widehat{\boldsymbol{h}})\boldsymbol{F}_M \quad (30)$$

after which the frequency-domain quantities

$$\underline{\widehat{\boldsymbol{X}}} \triangleq \boldsymbol{F}_M \widehat{\boldsymbol{X}} \quad (31)$$

$$\underline{\widehat{\boldsymbol{h}}} \triangleq \boldsymbol{F}_M^{1:L}\widehat{\boldsymbol{h}} \quad (32)$$

can be used to rewrite $\widehat{\boldsymbol{p}}$ as

$$\widehat{\boldsymbol{p}} = \mathrm{vec}\big(\sqrt{M}\boldsymbol{F}_M^\mathsf{H}\,\mathrm{Diag}(\underline{\widehat{\boldsymbol{h}}})\underline{\widehat{\boldsymbol{X}}}\big) - \overline{\nu}^\mathsf{p}\widehat{\boldsymbol{s}}. \quad (33)$$

Next we discuss PBiGAMP's nonlinear steps (R7)–(R8), which—according to (D1)—compute the posterior mean and variance of z_m given the likelihood function $p_{\mathsf{y}_m|\mathsf{z}_m}(y_m|z_m)$ from (16) and the prior $\mathsf{z}_m \sim \mathcal{CN}(\widehat{p}_m, \nu^\mathsf{p})$. Recall that the real and imaginary parts of $\mathcal{CN}(\widehat{p}_m, \nu^\mathsf{p})$ are independent Gaussian with means $\widehat{p}_m^{\,\mathrm{re}}$ and $\widehat{p}_m^{\,\mathrm{im}}$, respectively, and variance $\nu^\mathsf{p}/2$. Then, because the quantization $\mathcal{Q}(\cdot)$ is applied separately to real and imaginary components, we can separately compute the posterior means and variances for the real and imaginary components of z_m. Using $(g_{u-1}, g_u] \subset \mathbb{R}$ to denote the interval of u_m^re quantized to y_m^re, the posterior mean and variance of the real part of

z_m can be expressed as

$$\widehat{z}_m^{\text{re}} = \widehat{p}_m^{\text{re}} + \frac{\nu^{\text{p}}}{2}\frac{D_m^{\text{re}}}{E_m^{\text{re}}} \tag{34}$$

$$\nu_m^{\text{z,re}} = \frac{\nu^{\text{p}}}{2} + \frac{F_m^{\text{re}}}{E_m^{\text{re}}}\left(\frac{\nu^{\text{p}}}{2}\right)^2 - (\widehat{z}_m^{\text{re}} - \widehat{p}_m^{\text{re}})^2 \tag{35}$$

where

$$\begin{aligned} D_m^{\text{re}} &= \mathcal{N}\big(\widehat{p}_m^{\text{re}} - g_{u-1}; 0, (\sigma_w^2 + \nu^{\text{p}})/2\big) \\ &\quad - \mathcal{N}\big(\widehat{p}_m^{\text{re}} - g_u; 0, (\sigma_w^2 + \nu^{\text{p}})/2\big) \end{aligned} \tag{36}$$

$$E_m^{\text{re}} = \Phi\left(\frac{\widehat{p}_m^{\text{re}} - g_{u-1}}{\sqrt{(\sigma_w^2 + \nu^{\text{p}})/2}}\right) - \Phi\left(\frac{\widehat{p}_m^{\text{re}} - g_{u-1}}{\sqrt{(\sigma_w^2 + \nu^{\text{p}})/2}}\right) \tag{37}$$

$$\begin{aligned} F_m^{\text{re}} &= \frac{\widehat{p}_m^{\text{re}} - g_u}{(\sigma_w^2 + \nu^{\text{p}})/2}\mathcal{N}\big(\widehat{p}_m^{\text{re}} - g_u; 0, (\sigma_w^2 + \nu^{\text{p}})/2\big) \\ &\quad - \frac{\widehat{p}_m^{\text{re}} - g_{u-1}}{(\sigma_w^2 + \nu^{\text{p}})/2}\mathcal{N}\big(\widehat{p}_m^{\text{re}} - g_{u-1}; 0, (\sigma_w^2 + \nu^{\text{p}})/2\big). \end{aligned} \tag{38}$$

Similarly, the posterior mean and variance of the imaginary part of z_m can be computed using the same procedure, but with $\widehat{p}_m^{\text{im}}$ replacing $\widehat{p}_m^{\text{re}}$. Finally, for (R7)–(R8), the real and imaginary parts are combined as

$$\widehat{z}_m = \widehat{z}_m^{\text{re}} + \text{j}\widehat{z}_m^{\text{im}}, \quad \nu^{\text{z}} = \frac{1}{MK}\sum_{m=0}^{MK-1}\big(\nu_m^{\text{z,re}} + \nu_m^{\text{z,im}}\big). \tag{39}$$

Equations (34)–(38) can be derived following the procedures in [53, Chapter 3.9]; see [17, Appendix A] for further details.

Next we consider PBiGAMP steps (R11)–(R14). From (21)–(22), steps (R11) and (R13) become

$$\nu^{\text{r}} = \frac{1}{\nu^{\text{s}}\|\widehat{\boldsymbol{x}}\|^2} \tag{40}$$

$$\nu^{\text{q}} = \frac{1}{\nu^{\text{s}}\|\widehat{\boldsymbol{h}}\|^2}. \tag{41}$$

For step (R12), we use (22) and (26) to write

$$\widehat{r}_l = \widehat{h}_l + \nu^{\text{r}}\widehat{\boldsymbol{z}}^{(*,l)\mathsf{H}}\widehat{\boldsymbol{s}} - \nu^{\text{r}}\nu^{\text{s}}\nu^{\text{x}}\widehat{h}_l\sum_{n=0}^{MK-1}\|\boldsymbol{z}^{(n,l)}\|^2 \tag{42}$$

$$= \widehat{h}_l(1 - MK\nu^{\text{r}}\nu^{\text{s}}\nu^{\text{x}}) + \nu^{\text{r}}\text{vec}(\boldsymbol{J}_l\widehat{\boldsymbol{X}})^{\mathsf{H}}\text{vec}(\widehat{\boldsymbol{S}}) \tag{43}$$

$$= \widehat{h}_l(1 - MK\nu^{\text{r}}\nu^{\text{s}}\nu^{\text{x}}) + \nu^{\text{r}}\sum_{k=1}^{K}(\boldsymbol{J}_l\widehat{\boldsymbol{x}}_k)^{\mathsf{H}}\widehat{\boldsymbol{s}}_k, \tag{44}$$

where $\widehat{\boldsymbol{S}} \in \mathbb{C}^{M\times K}$ is a reshaping of $\widehat{\boldsymbol{s}}$ and where $\widehat{\boldsymbol{x}}_k$ and $\widehat{\boldsymbol{s}}_k$ are the kth columns of $\widehat{\boldsymbol{X}}$ and $\widehat{\boldsymbol{S}}$. Thus $\widehat{\boldsymbol{r}} \triangleq [\widehat{r}_0, \ldots, \widehat{r}_{L-1}]^{\mathsf{T}}$ takes the form

$$\widehat{\boldsymbol{r}} = \widehat{\boldsymbol{h}}(1 - MK\nu^{\text{r}}\nu^{\text{x}}\nu^{\text{s}}) + \nu^{\text{r}}\sum_{k=1}^{K}\big[\boldsymbol{J}_0\widehat{\boldsymbol{x}}_k, \ldots, \boldsymbol{J}_{L-1}\widehat{\boldsymbol{x}}_k\big]^{\mathsf{H}}\widehat{\boldsymbol{s}}_k. \tag{45}$$

Since $\big[\boldsymbol{J}_0\widehat{\boldsymbol{x}}_k, \ldots, \boldsymbol{J}_{L-1}\widehat{\boldsymbol{x}}_k\big]$ are the first L columns of the circulant matrix with first column $\widehat{\boldsymbol{x}}_k$, (30) implies

$$\big[\boldsymbol{J}_0\widehat{\boldsymbol{x}}_k, \ldots, \boldsymbol{J}_{L-1}\widehat{\boldsymbol{x}}_k\big] = \sqrt{M}\boldsymbol{F}_M^{\mathsf{H}}\text{Diag}(\boldsymbol{F}_M\widehat{\boldsymbol{x}}_k)\boldsymbol{F}_M^{1:L}. \tag{46}$$

Plugging (46) into (45), and defining $\underline{\widehat{\boldsymbol{x}}}_k \triangleq \boldsymbol{F}_M\widehat{\boldsymbol{x}}_k$ (i.e., the kth column of $\widehat{\underline{\boldsymbol{X}}}$) and $\underline{\widehat{\boldsymbol{s}}}_k \triangleq \boldsymbol{F}_M\widehat{\boldsymbol{s}}_k$, we get

$$\widehat{\boldsymbol{r}} = \widehat{\boldsymbol{h}}(1 - MK\nu^{\text{r}}\nu^{\text{x}}\nu^{\text{s}}) + \sqrt{M}\nu^{\text{r}}(\boldsymbol{F}_M^{1:L})^{\mathsf{H}}\sum_{k=1}^{K}\underline{\widehat{\boldsymbol{x}}}_k^* \odot \underline{\boldsymbol{s}}_k. \tag{47}$$

A similar derivation reduces PBiGAMP step (R14) to

$$\widehat{\boldsymbol{q}} = \widehat{\boldsymbol{x}}(1 - L\nu^{\text{q}}\nu^{\text{h}}\nu^{\text{s}}) + \sqrt{M}\nu^{\text{q}}\text{vec}\big(\boldsymbol{F}_M^{\mathsf{H}}\text{Diag}(\underline{\widehat{\boldsymbol{h}}})^{\mathsf{H}}\underline{\widehat{\boldsymbol{S}}}\big), \tag{48}$$

where $\underline{\widehat{\boldsymbol{S}}} \triangleq \boldsymbol{F}_M\widehat{\boldsymbol{S}}$.

Next we consider PBiGAMP steps (R15)–(R16), which—according to (D2)—compute the posterior mean and variance of h_l given the GMM prior (8) and the likelihood function $\mathcal{CN}(\widehat{r}_l; h_l, \nu^{\text{r}})$. From [40], the posterior is

$$p_{\mathsf{h}_l|\mathsf{r}_l}(h_l \,|\, \widehat{r}_l; \nu_l^{\text{r}}) = \sum_{d=1}^{D}\overline{\lambda}_{l,d}\mathcal{CN}\left(h_l; \frac{\nu_{l,d}\widehat{r}_l}{\nu_{l,d} + \nu_l^{\text{r}}}, \frac{\nu_{l,d}\nu_l^{\text{r}}}{\nu_{l,d} + \nu_l^{\text{r}}}\right) \tag{49}$$

$$\overline{\lambda}_{l,d} = \frac{\lambda_{l,d}\mathcal{CN}(\widehat{r}_l; 0, \nu_{l,d} + \nu_l^{\text{r}})}{\sum_{d'=1}^{D}\lambda_{l,d'}\mathcal{CN}(\widehat{r}_l; 0, \nu_{l,d'} + \nu_l^{\text{r}})}, \tag{50}$$

which is also a GMM. The corresponding mean and variance follow straightforwardly as

$$\widehat{h}_l = \sum_{d=1}^{D}\overline{\lambda}_{l,d}\frac{\nu_{l,d}\widehat{r}_l}{\nu_{l,d} + \nu_l^{\text{r}}} \tag{51}$$

$$\nu_l^{\text{h}} = \sum_{d=1}^{D}\overline{\lambda}_{l,d}\left(\frac{\nu_{l,d}\nu_l^{\text{r}}}{\nu_{l,d} + \nu_l^{\text{r}}} + \left|\frac{\nu_{l,d}\widehat{r}_l}{\nu_{l,d} + \nu_l^{\text{r}}}\right|^2\right) - |\widehat{h}_l|^2. \tag{52}$$

Finally, we consider PBiGAMP steps (R17)–(R18), which—according to (D3)—compute the posterior mean and variance of x_n given the discrete symbol prior (20) and the likelihood function $\mathcal{CN}(\widehat{q}_n; x_n, \nu^{\text{q}})$. In this case, the posterior is

$$p_{\mathsf{x}_n|\mathsf{q}_n}(x_n \,|\, \widehat{q}_n; \nu_n^{\text{q}}) = \sum_{j=1}^{2^A}\overline{\gamma}_{n,j}\delta(x_n - s^{(j)}) \tag{53}$$

$$\overline{\gamma}_{n,j} = \frac{\Pr\{\mathsf{x}_n = s^{(j)}\}\mathcal{CN}\big(s^{(j)}; \widehat{q}_n, \nu_n^{\text{q}}\big)}{\sum_{j'=1}^{2^A}\Pr\{\mathsf{x}_n = s^{(j')}\}\mathcal{CN}\big(s^{(j')}; \widehat{q}_n, \nu_n^{\text{q}}\big)}, \tag{54}$$

which is a discrete distribution with support on $\mathcal{S}$. The posterior mean and variance follow as

$$\widehat{x}_n = \sum_{j=1}^{2^A}\overline{\gamma}_{n,j}s^{(j)} \tag{55}$$

$$\nu_n^{\text{x}} = \sum_{j=1}^{2^A}\overline{\gamma}_{n,j}|s^{(j)} - \widehat{x}_n|^2. \tag{56}$$

TABLE II
SOFT EQUALIZATION VIA SCALAR-VARIANCE PBIGAMP

Definitions:

$$p_{\mathsf{z}_m|\mathsf{p}_m}(z\,|\,\widehat{p};\nu^{\mathsf{p}}) \triangleq \frac{p_{\mathsf{y}_m|\mathsf{z}_m}(y_m\,|\,z)\,\mathcal{CN}(z;\widehat{p},\nu^{\mathsf{p}})}{\int p_{\mathsf{y}_m|\mathsf{z}_m}(y_m\,|\,z')\,\mathcal{CN}(z';\widehat{p},\nu^{\mathsf{p}})\,\mathrm{d}z'} \quad \text{(D1)}$$

$$p_{\mathsf{h}_l|\mathsf{r}_l}(h\,|\,\widehat{r};\nu^{\mathsf{r}}) \triangleq \frac{p_{\mathsf{h}_l}(h)\,\mathcal{CN}(\widehat{r};h,\nu^{\mathsf{r}})}{\int p_{\mathsf{h}_l}(h')\,\mathcal{CN}(\widehat{r};h',\nu^{\mathsf{r}})\,\mathrm{d}h'} \quad \text{(D2)}$$

$$p_{\mathsf{x}_n|\mathsf{q}_n}(x\,|\,\widehat{q};\nu^{\mathsf{q}}) \triangleq \frac{p_{\mathsf{x}_n}(x)\,\mathcal{CN}(\widehat{q};x,\nu^{\mathsf{q}})}{\int p_{\mathsf{x}_n}(x')\,\mathcal{CN}(\widehat{q};x',\nu^{\mathsf{q}})\,\mathrm{d}x'} \quad \text{(D3)}$$

Initialization:

$$\boldsymbol{x}_{0\mathrm{G}} = [\boldsymbol{0}_{N_{\mathrm{D}}}^{\mathsf{T}}, \boldsymbol{x}_{\mathrm{G}}^{\mathsf{T}}]^{\mathsf{T}}$$

$$\widehat{\boldsymbol{X}}[1] = \big[\boldsymbol{x}_{\mathrm{P},1}, \ldots, \boldsymbol{x}_{\mathrm{P},K_{\mathrm{P}}}, \boldsymbol{x}_{0\mathrm{G}}, \ldots, \boldsymbol{x}_{0\mathrm{G}}\big],\ \nu^{\mathsf{x}}[1] = \tfrac{K_{\mathrm{D}}N_{\mathrm{D}}}{MK}$$

$$\widehat{\boldsymbol{h}}[1] = \widehat{\boldsymbol{h}}_{\text{init}},\ \nu^{\mathsf{h}}[1] = \nu^{\mathsf{h}}_{\text{init}},\ \widehat{\boldsymbol{S}}[0] = \boldsymbol{0}_{M\times K}$$

For $t = 1, \ldots T_{\max}$

$$\underline{\widehat{\boldsymbol{X}}}[t] = \boldsymbol{F}_M \widehat{\boldsymbol{X}}[t] \quad \text{(E1)}$$

$$\underline{\widehat{\boldsymbol{h}}}[t] = \boldsymbol{F}_M^{1:L} \widehat{\boldsymbol{h}}[t] \quad \text{(E2)}$$

$$\overline{\nu}^{\mathsf{p}}[t] = \nu^{\mathsf{x}}[t]\big\|\widehat{\boldsymbol{h}}[t]\big\|^2 + \tfrac{L}{MK}\nu^{\mathsf{h}}[t]\big\|\widehat{\boldsymbol{X}}[t]\big\|_F^2 \quad \text{(E3)}$$

$$\nu^{\mathsf{p}}[t] = \overline{\nu}^{\mathsf{p}}[t] + L\nu^{\mathsf{h}}[t]\nu^{\mathsf{x}}[t] \quad \text{(E4)}$$

$$\widehat{\boldsymbol{P}}[t] = \sqrt{M}\boldsymbol{F}_M^{\mathsf{H}}\mathrm{Diag}(\underline{\widehat{\boldsymbol{h}}}[t])\underline{\widehat{\boldsymbol{X}}}[t] - \overline{\nu}^{\mathsf{p}}[t]\widehat{\boldsymbol{S}}[t-1] \quad \text{(E5)}$$

$$\nu^{\mathsf{z}}[t] = \tfrac{1}{MK}\textstyle\sum_{m=1}^{M-1}\sum_{k=1}^{K}\mathrm{var}\{\mathsf{z}_{mk}\,|\,\widehat{p}_{mk}[t];\nu^{\mathsf{p}}[t]\} \quad \text{(E6)}$$

$$\forall m,k\colon \widehat{z}_{mk}[t] = \mathbb{E}[\mathsf{z}_{mk}\,|\,\mathsf{p}_{mk}=\widehat{p}_{mk}[t];\nu^{\mathsf{p}}[t]] \quad \text{(E7)}$$

$$\nu^{\mathsf{s}}[t] = \big(1-\nu^{\mathsf{z}}[t]/\nu^{\mathsf{p}}[t]\big)/\nu^{\mathsf{p}}[t] \quad \text{(E8)}$$

$$\widehat{\boldsymbol{S}}[t] = \big(\widehat{\boldsymbol{Z}}[t] - \widehat{\boldsymbol{P}}[t]\big)/\nu^{\mathsf{p}}[t] \quad \text{(E9)}$$

$$\underline{\widehat{\boldsymbol{S}}}[t] = \boldsymbol{F}_M\widehat{\boldsymbol{S}}[t] \quad \text{(E10)}$$

$$\nu^{\mathsf{r}}[t] = \big(\nu^{\mathsf{s}}[t]\big\|\widehat{\boldsymbol{X}}[t]\big\|_F^2\big)^{-1} \quad \text{(E11)}$$

$$\widehat{\boldsymbol{r}}[t] = \nu^{\mathsf{r}}[t]\sqrt{M}(\boldsymbol{F}_M^{1:L})^{\mathsf{H}}\big(\underline{\widehat{\boldsymbol{X}}}[t]^* \odot \underline{\widehat{\boldsymbol{S}}}[t]\big)\boldsymbol{1}_K + \big(1 - MK\nu^{\mathsf{r}}[t]\nu^{\mathsf{x}}[t]\nu^{\mathsf{s}}[t]\big)\widehat{\boldsymbol{h}}[t] \quad \text{(E12)}$$

$$\nu^{\mathsf{q}}[t] = \big(\nu^{\mathsf{s}}[t]\big\|\widehat{\boldsymbol{h}}[t]\big\|^2\big)^{-1} \quad \text{(E13)}$$

$$\widehat{\boldsymbol{Q}}[t] = \sqrt{M}\nu^{\mathsf{q}}[t]\boldsymbol{F}_M^{\mathsf{H}}\mathrm{Diag}(\underline{\widehat{\boldsymbol{h}}}[t])^{\mathsf{H}}\underline{\widehat{\boldsymbol{S}}}[t] + \big(1 - L\nu^{\mathsf{q}}[t]\nu^{\mathsf{h}}[t]\nu^{\mathsf{s}}[t]\big)\widehat{\boldsymbol{X}}[t] \quad \text{(E14)}$$

$$\nu^{\mathsf{h}}[t+1] = \tfrac{1}{L}\textstyle\sum_{l=0}^{L-1}\mathrm{var}\{\mathsf{h}_l\,|\,\mathsf{r}_l=\widehat{r}_l[t];\nu^{\mathsf{r}}[t]\} \quad \text{(E15)}$$

$$\forall l\colon \widehat{h}_l[t+1] = \mathbb{E}[\mathsf{h}_l\,|\,\mathsf{r}_l=\widehat{r}_l[t];\nu^{\mathsf{r}}[t]] \quad \text{(E16)}$$

$$\nu^{\mathsf{x}}[t+1] = \tfrac{1}{MK}\textstyle\sum_{m=0}^{M-1}\sum_{k=1}^{K}\mathrm{var}\{\mathsf{x}_{mk}\,|\,\widehat{q}_{mk}[t];\nu^{\mathsf{q}}[t]\} \quad \text{(E17)}$$

$$\forall m,k\colon \widehat{x}_{mk}[t+1] = \mathbb{E}[\mathsf{x}_{mk}\,|\,\mathsf{q}_{mk}=\widehat{q}_{mk}[t];\nu^{\mathsf{q}}[t]] \quad \text{(E18)}$$

end

Note that $\{\overline{\gamma}_{n,j}\}_{j=1}^{2^A}$ is the posterior pmf on x_n. It can be converted to posterior pmfs on the coded bits $\{\mathsf{c}_{n,a}\}_{a=1}^{A}$ via

$$\Pr\{\mathsf{c}_{n,a}=1\,|\,\widehat{q}_n\} = \sum_{j=1\ldots 2^A\,|\,c_a^{(j)}=1} \Pr\{\mathsf{c}_n=\boldsymbol{c}^{(j)}\,|\,\widehat{q}_n\} \quad (57)$$

$$= \sum_{\substack{j=1\ldots 2^A\\ c_a^{(j)}=1}} \sum_{j'=1}^{2^A} \underbrace{\Pr\{\mathsf{c}_n=\boldsymbol{c}^{(j)}|\mathsf{x}_n=s^{(j')}\}}_{\delta_{j-j'}} \underbrace{\Pr\{\mathsf{x}_n=s^{(j')}\,|\,\widehat{q}_n\}}_{\overline{\gamma}_{n,j'}} \quad (58)$$

$$= \sum_{j=1\ldots 2^A\,|\,c_a^{(j)}=1} \overline{\gamma}_{n,j}. \quad (59)$$

The PBiGAMP-based soft equalization procedure is summarized in Table II using $(M\times K)$-matricized versions of $\widehat{\boldsymbol{p}}$, $\widehat{\boldsymbol{q}}$, and $\widehat{\boldsymbol{x}}$ denoted by $\widehat{\boldsymbol{P}}$, $\widehat{\boldsymbol{Q}}$, and $\widehat{\boldsymbol{X}}$, respectively. Its complexity is dominated by the $4K+2$ DFT-matrix multiplies in steps (E1), (E2), (E5), (E10), (E12), and (E14), which consume a total of $O(MK\log M)$ operations per iteration, or $O(\log M)$ operations per symbol per iteration, when an FFT is used. All other lines in Table II consume a total of $O(MK)$ operations per iteration, or $O(1)$ operations per symbol per iteration.

For notational simplicity, the table does not reflect the fact that the first K_{P} columns of $\widehat{\boldsymbol{X}}$ are known pilots and the last N_{G} elements of the remaining columns in $\widehat{\boldsymbol{X}}$ are known guards. For those known elements, the mean and variance computations in (E17)–(E18) can be omitted. Likewise, there is no need to compute the first K_{P} columns of $\underline{\widehat{\boldsymbol{X}}}$ in (E1) or the first K_{P} columns of $\widehat{\boldsymbol{Q}}$ in (E14), reducing the number of required FFTs by $2K_{\mathrm{P}}$.

D. Turbo Equalization

As described in Section III-A, we would like to compute (approximate) posterior marginal bit probabilities $\{p(b_i|\boldsymbol{y})\}_{i=1}^{N_{\mathrm{b}}}$ using the SPA, which is the usual approach to turbo equalization [36]. Because exact SPA is intractable for the soft-equalization subgraph in Fig. 3, we use the PBiGAMP approximation, as described in Section III-B, on that subgraph. We now detail the remaining steps in the SPA, for completeness.

Roughly speaking, messages are passed on the factor graph in Fig. 3 from the left to the right and back again. One such forward-backward pass will be referred to as a turbo iteration. During a single turbo iteration, soft equalization using PBiGAMP is alternated with soft decoding using a standard decoder/interleaver. The SPA dictates that "extrinsic" information is passed between nodes on the graph and hence between the subgraphs in Fig. 3. For a discrete random variable, the extrinsic message is a pmf formed by dividing the posterior pmf by the prior pmf. Additional details are given below.

During each turbo iteration, extrinsic information on the coded bits $\mathsf{c}_{n,a}$ is passed from the soft decoder to PBiGAMP, where it is treated as prior information in (20) to determine the symbol priors $\gamma_{n,j}$. PBiGAMP is then run to convergence, generating the symbol posteriors $\overline{\gamma}_{n,j}$. The symbol posteriors are used in (59) to determine the coded-bit posteriors, which are then converted to extrinsic form and passed to the soft decoder. The soft decoder accepts this extrinsic information from PBiGAMP, treating it as a prior on the coded bits. It then computes posteriors on the coded bits, converts them to extrinsic form, and passes them to PBiGAMP for the next turbo iteration.

E. Learning the Channel Prior

The GMM prior (8) requires specification of the weights and variances $\{\boldsymbol{\lambda}_l, \boldsymbol{\nu}_l\}_{l=0}^{L-1}$. In the simple case where the coefficients are modeled as identically distributed, the set $\{\boldsymbol{\lambda}_l, \boldsymbol{\nu}_l\}_{l=0}^{L-1}$ reduces to the pair $\boldsymbol{\lambda}, \boldsymbol{\nu}$. The "EM-GM-AMP" paper [40] showed how this pair can be learned from the observations $\boldsymbol{y}$ using a combination of EM and AMP, and [37] showed how EM can be combined with PBiGAMP in a similar manner. In Section V, we investigate the performance of this EM-GM-PBiGAMP method on the channels described in Section II-C using GMM order $D=2$. More generally, one could partition the coefficients $\{\mathsf{h}_l\}_{l=0}^{L-1}$ into subsets and learn a different weight and variance for each subset, as discussed in [39]. Typically, the EM update is performed in line (E16) once per PBiGAMP iteration, so that the computational burden of EM is very minor.

F. Scaling the Channel Estimate

With few-bit ADCs, channel amplitude information is degraded due to quantization (and completely lost in the case of a one-bit ADC). Thus, we find that channel-estimation performance can be improved by appropriately scaling the channel estimate. To do this, we exploit the fact that

$$\mathbb{E}[\|\boldsymbol{u}\|^2 \,|\, \boldsymbol{h}] = \mathrm{tr}\{\mathbb{E}[\boldsymbol{u}\boldsymbol{u}^{\mathsf{H}} \,|\, \boldsymbol{h}]\} \tag{60}$$

$$= \mathrm{tr}\{(\boldsymbol{I}_K \otimes \boldsymbol{H})\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^{\mathsf{H}}](\boldsymbol{I}_K \otimes \boldsymbol{H})^{\mathsf{H}}\} + MK\sigma_w^2 \tag{61}$$

$$= \sigma_x^2 \mathrm{tr}\{\boldsymbol{I}_K \otimes \boldsymbol{H}\boldsymbol{H}^{\mathsf{H}}\} + MK\sigma_w^2 \tag{62}$$

$$= K\sigma_x^2 \mathrm{tr}\{\boldsymbol{H}\boldsymbol{H}^{\mathsf{H}}\} + MK\sigma_w^2 \tag{63}$$

$$= MK\sigma_x^2\|\boldsymbol{h}\|^2 + MK\sigma_w^2 \tag{64}$$

due to the circulant nature of $\boldsymbol{H}$, and so

$$\|\boldsymbol{h}\| = \sqrt{\frac{\mathbb{E}[\|\boldsymbol{u}\|^2 \,|\, \boldsymbol{h}]/(MK) - \sigma_w^2}{\sigma_x^2}}. \tag{65}$$

Assuming that the average received-signal power $\mathbb{E}[\|\boldsymbol{u}\|^2 \,|\, \boldsymbol{h}]/(MK)$ can be measured[2] prior to the ADC (as is typically done as part of automatic gain control), the true channel norm can be computed from (65) and the channel estimate $\widehat{\boldsymbol{h}}$ can be scaled so that its norm matches the true one. We note that a similar technique was used in [24]. With PBiGAMP, we scale the output of line (E16) in this manner at each iteration.

IV. Benchmark Methods

We now describe two methods that will be used later for performance evaluation: PBiGAMP with Bussgang linearization, and pilot-aided channel estimation plus LMMSE decoding.

A. PBiGAMP With Bussgang Linearization

The PBiGAMP method proposed in Section III uses a non-Gaussian likelihood function $p_{\mathsf{y}_m|\mathsf{z}_m}$ that results directly from the quantization model (5). An alternative explored in the literature is the use of an AWGN approximation of $p_{\mathsf{y}_m|\mathsf{z}_m}$ based on a Bussgang linearization [54]. This leads to a simplified approach that tends to perform well under mild quantization. We briefly summarize the Bussgang approach below.[3]

The Bussgang linearization first writes the nonlinear quantization operation $\boldsymbol{y} = \mathcal{Q}(\boldsymbol{u})$ as

$$\boldsymbol{y} = \boldsymbol{G}_{\mathbf{y}}\boldsymbol{u} + \boldsymbol{e}, \tag{66}$$

where $\boldsymbol{G}_{\mathbf{y}}$ is the LMMSE estimator of $\boldsymbol{y}$ from $\boldsymbol{u}$, i.e.,

$$\boldsymbol{G}_{\mathbf{y}} = \mathbb{E}[\boldsymbol{y}\boldsymbol{u}^{\mathsf{H}}]\mathbb{E}[\boldsymbol{u}\boldsymbol{u}^{\mathsf{H}}]^{-1}, \tag{67}$$

and $\boldsymbol{e} \triangleq \boldsymbol{y} - \boldsymbol{G}_{\mathbf{y}}\boldsymbol{u}$ is the estimation error. Due to the orthogonality principle, we know that $\mathbb{E}[\boldsymbol{u}\boldsymbol{e}^{\mathsf{H}}] = \boldsymbol{0}$, i.e., the Bussgang error $\boldsymbol{e}$ is uncorrelated with the quantizer input $\boldsymbol{u}$.

Plugging the expression for $\boldsymbol{u}$ from (3) into (66), we get

$$\boldsymbol{y} = \boldsymbol{G}_{\mathbf{y}}(\boldsymbol{I}_K \otimes \boldsymbol{H})\boldsymbol{x} + \underbrace{\boldsymbol{G}_{\mathbf{y}}\boldsymbol{w} + \boldsymbol{e}}_{\triangleq\, \widetilde{\boldsymbol{w}}}, \tag{68}$$

where we can interpret $\boldsymbol{G}_{\mathbf{y}}(\boldsymbol{I}_K \otimes \boldsymbol{H})$ as the effective channel and $\widetilde{\boldsymbol{w}}$ as the effective noise. Although non-Gaussian, $\widetilde{\boldsymbol{w}}$ is approximately uncorrelated with the signal $\boldsymbol{x}$, in that

$$\mathbb{E}[\boldsymbol{x}\widetilde{\boldsymbol{w}}^{\mathsf{H}}] = \mathbb{E}[\boldsymbol{x}\boldsymbol{w}^{\mathsf{H}}]\boldsymbol{G}_{\mathbf{y}}^{\mathsf{H}} + \mathbb{E}[\boldsymbol{x}\boldsymbol{e}^{\mathsf{H}}] \tag{69}$$

$$= \mathbb{E}[\boldsymbol{x}\boldsymbol{e}^{\mathsf{H}}] \tag{70}$$

$$= \mathbb{E}\big\{\mathbb{E}[\boldsymbol{x}\boldsymbol{e}^{\mathsf{H}}|\boldsymbol{u}]\big\} = \mathbb{E}\big\{\mathbb{E}[\boldsymbol{x}|\boldsymbol{u}]\boldsymbol{e}^{\mathsf{H}}\big\} \tag{71}$$

$$\approx \mathbb{E}[\boldsymbol{G}_{\mathbf{x}}\boldsymbol{u}\boldsymbol{e}^{\mathsf{H}}] = \boldsymbol{G}_{\mathbf{x}}\mathbb{E}[\boldsymbol{u}\boldsymbol{e}^{\mathsf{H}}] \tag{72}$$

$$= \boldsymbol{0}, \tag{73}$$

where (70) follows from $\mathbb{E}[\boldsymbol{x}\boldsymbol{w}^{\mathsf{H}}] = \boldsymbol{0}$, (71) follows from the fact that $\boldsymbol{e} = \mathcal{Q}(\boldsymbol{u}) - \boldsymbol{G}_{\mathbf{y}}\boldsymbol{u}$ is deterministic when conditioned on $\boldsymbol{u}$, and (72) approximates $\mathbb{E}[\boldsymbol{x}|\boldsymbol{u}]$ by the LMMSE estimate $\boldsymbol{G}_{\mathbf{x}}\boldsymbol{u}$ of $\boldsymbol{x}$ from $\boldsymbol{u}$. This approximation becomes exact when $\boldsymbol{x}$ and $\boldsymbol{u}$ are jointly Gaussian. Finally, equation (73) follows from $\mathbb{E}[\boldsymbol{u}\boldsymbol{e}^{\mathsf{H}}] = \boldsymbol{0}$.

Note that $\boldsymbol{w}$ and $\boldsymbol{e}$ are also uncorrelated, in that

$$\mathbb{E}[\boldsymbol{w}\boldsymbol{e}^{\mathsf{H}}] = \mathbb{E}\big[\mathbb{E}[\boldsymbol{w}\boldsymbol{e}^{\mathsf{H}}|\boldsymbol{u}]\big] \tag{74}$$

$$= \mathbb{E}\big[\mathbb{E}[\boldsymbol{w}|\boldsymbol{u}]\boldsymbol{e}^{\mathsf{H}}\big] \tag{75}$$

$$= \mathbb{E}[\boldsymbol{G}_{\mathbf{w}}\boldsymbol{u}\boldsymbol{e}^{\mathsf{H}}] = \boldsymbol{G}_{\mathbf{w}}\mathbb{E}[\boldsymbol{u}\boldsymbol{e}^{\mathsf{H}}] \tag{76}$$

$$= \boldsymbol{0}, \tag{77}$$

where (75) results because $\boldsymbol{e}$ is deterministic conditioned on $\boldsymbol{u}$, (76) results because $\boldsymbol{w}$ and $\boldsymbol{u}$ are jointly Gaussian, with $\boldsymbol{G}_{\mathbf{w}}$ denoting the LMMSE estimator of $\boldsymbol{w}$ from $\boldsymbol{u}$, and (77) follows from $\mathbb{E}[\boldsymbol{u}\boldsymbol{e}^{\mathsf{H}}] = \boldsymbol{0}$. As a consequence of (77), the covariance of $\widetilde{\boldsymbol{w}}$ reduces to

$$\mathbb{E}[\widetilde{\boldsymbol{w}}\widetilde{\boldsymbol{w}}^{\mathsf{H}}] = \sigma_w^2\boldsymbol{G}_{\mathbf{y}}\boldsymbol{G}_{\mathbf{y}}^{\mathsf{H}} + \mathbb{E}[\boldsymbol{e}\boldsymbol{e}^{\mathsf{H}}]. \tag{78}$$

For uniform quantization with MMSE stepsize Δ_b [55] (recall (5)), the LMMSE matrix $\boldsymbol{G}_{\mathbf{y}}$ has a simple form. To see this, we first define the quantization error

$$\boldsymbol{q} \triangleq \boldsymbol{y} - \boldsymbol{u}. \tag{79}$$

Note, from (3) and the fact that $\boldsymbol{H}$ is circulant with first column $\boldsymbol{h}$, that $u_m = \sum_{l=0}^{M-1} h_{\langle m-l\rangle_M}\, x_{\lfloor m/M\rfloor M+l}$, where $\langle n\rangle_M$ denotes n-modulo-M. Thus, if we treat the components of $\boldsymbol{x}$ as i.i.d., then the components of $\boldsymbol{u}$ will be identically distributed. Consequently, the components of $\boldsymbol{y} = \mathcal{Q}(\boldsymbol{u})$ will be identically distributed, as will those of $\boldsymbol{q}$. In this case, the results in [54] imply

$$\mathbb{E}[\boldsymbol{u}\boldsymbol{q}^{\mathsf{H}}] = -\eta\mathbb{E}[\boldsymbol{u}\boldsymbol{u}^{\mathsf{H}}] = \mathbb{E}[\boldsymbol{q}\boldsymbol{u}^{\mathsf{H}}] \tag{80}$$

$$\mathbb{E}[\boldsymbol{q}\boldsymbol{q}^{\mathsf{H}}] \approx \eta\mathbb{E}[\boldsymbol{u}\boldsymbol{u}^{\mathsf{H}}] - (1-\eta)\eta\,\mathrm{Nondiag}(\mathbb{E}[\boldsymbol{u}\boldsymbol{u}^{\mathsf{H}}]) \tag{81}$$

$$= \eta^2\mathbb{E}[\boldsymbol{u}\boldsymbol{u}^{\mathsf{H}}] + (1-\eta)\eta\mathrm{Diag}(\mathrm{diag}(\mathbb{E}[\boldsymbol{u}\boldsymbol{u}^{\mathsf{H}}])), \tag{82}$$

[2]To measure the average received-signal power, it suffices to use an ADC with a relatively low sampling rate, which is inexpensive in both cost and power consumption.

[3]Our summary includes an explanation of why the effective noise $\widetilde{\boldsymbol{w}}$ is uncorrelated with the signal $\boldsymbol{x}$, which is missing from [54], as well as specializations relevant to (3).

where

$$\eta \triangleq \frac{\mathbb{E}[|q_m|^2]}{\mathbb{E}[|u_m|^2]}. \tag{83}$$

The approximation (81) would be exact if q_m and $y_{m'}$ were jointly Gaussian for all $m \neq m'$. From (67), we now see that

$$\boldsymbol{G}_{\mathrm{y}} = \mathbb{E}[(\boldsymbol{u}+\boldsymbol{q})\boldsymbol{u}^{\mathsf{H}}]\mathbb{E}[\boldsymbol{u}\boldsymbol{u}^{\mathsf{H}}]^{-1} \tag{84}$$

$$= (1-\eta)\boldsymbol{I}, \tag{85}$$

where (85) follows from (80).

We can now compute the effective noise covariance (78). Noting from (66), (79), and (85) that

$$\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{G}_{\mathrm{y}}\boldsymbol{u} = \boldsymbol{u} + \boldsymbol{q} - (1-\eta)\boldsymbol{u} = \eta\boldsymbol{u} + \boldsymbol{q}, \tag{86}$$

we have

$$\mathbb{E}[\boldsymbol{e}\boldsymbol{e}^{\mathsf{H}}] = \mathbb{E}[(\eta\boldsymbol{u}+\boldsymbol{q})(\eta\boldsymbol{u}+\boldsymbol{q})^{\mathsf{H}}] \tag{87}$$

$$= \eta^2\mathbb{E}[\boldsymbol{u}\boldsymbol{u}^{\mathsf{H}}] + \eta\mathbb{E}[\boldsymbol{u}\boldsymbol{q}^{\mathsf{H}}] + \eta\mathbb{E}[\boldsymbol{q}\boldsymbol{u}^{\mathsf{H}}] + \mathbb{E}[\boldsymbol{q}\boldsymbol{q}^{\mathsf{H}}] \tag{88}$$

$$= \mathbb{E}[\boldsymbol{q}\boldsymbol{q}^{\mathsf{H}}] - \eta^2\mathbb{E}[\boldsymbol{u}\boldsymbol{u}^{\mathsf{H}}] \tag{89}$$

$$= (1-\eta)\eta\mathrm{Diag}(\mathrm{diag}(\mathbb{E}[\boldsymbol{u}\boldsymbol{u}^{\mathsf{H}}])), \tag{90}$$

where (89) follows from (80) and (90) follows from (82). Since

$$\mathbb{E}[|u_m|^2] = \mathbb{E}\big\{[\boldsymbol{I}\otimes\boldsymbol{H}]_{m,:}\boldsymbol{x}\boldsymbol{x}^{\mathsf{H}}[\boldsymbol{I}\otimes\boldsymbol{H}]_{m,:}^{\mathsf{H}}\big\} + \sigma_w^2 \tag{91}$$

$$= \sigma_x^2\mathbb{E}[\|\boldsymbol{h}\|^2] + \sigma_w^2, \tag{92}$$

equations (78), (85), (90), and (92) imply

$$\mathbb{E}[\widetilde{\boldsymbol{w}}\widetilde{\boldsymbol{w}}^{\mathsf{H}}] = (1-\eta)\eta(\sigma_x^2\mathbb{E}\{\|\boldsymbol{h}\|^2\} + \sigma_w^2)\boldsymbol{I} + (1-\eta)^2\sigma_w^2\boldsymbol{I} \tag{93}$$

$$= \underbrace{(1-\eta)(\eta\sigma_x^2\mathbb{E}\{\|\boldsymbol{h}\|^2\} + \sigma_w^2)}_{\triangleq\,\sigma_{\widetilde{w}}^2}\boldsymbol{I}. \tag{94}$$

Note that, in practice, $\mathbb{E}[|u_m|^2]$ can be estimated by measuring the input power to the ADC.

Finally, plugging (85) into (68), we get

$$\boldsymbol{y} = (1-\eta)(\boldsymbol{I}_K\otimes\boldsymbol{H})\boldsymbol{x} + \widetilde{\boldsymbol{w}}. \tag{95}$$

For the Bussgang approximation, we use (95), while treating the non-Gaussian effective noise $\widetilde{\boldsymbol{w}}$ as if it was AWGN with variance $\sigma_{\widetilde{w}}^2$ from (94).

In going from standard to Bussgang-linearized PBiGAMP, changes manifest only in lines (R7)–(R8) of Table I. In either case, the complexity of lines (R7)–(R8) is $O(MK)$ operations per frame, or $O(1)$ operations per symbol, recalling the discussion at the end of Section III-C. So, like PBiGAMP, the complexity of Bussgang-linearized PBiGAMP is $O(\log M)$ operations per symbol.

B. Pilot-Aided Channel Estimation and LMMSE Decoding

A computationally simpler benchmark is as follows. First, using the standard correlation-based approach that leverages the perfect aperiodic autocorrelation property of Golay sequences described in [56, Section 7.3.3.1], we obtain $\widehat{\boldsymbol{H}}$. Next, treating the channel estimate as if it were perfect, we perform linear-MMSE (LMMSE) turbo decoding on the Bussgang-linearized model (95). Details on the latter are provided below.

For each turbo iteration, we first convert the extrinsic information output by the coder into the data-symbol pmfs $\gamma_{n,j}$ via (20), and then we convert these pmfs into the prior symbol mean and variance vectors $\boldsymbol{\mu}$ and $\boldsymbol{v}$ via (55)–(56). At the very first turbo iteration, however, we set $\mu_n = 0$ and $v_n = 1$ for data indices n (assuming unit-variance symbols) and $\mu_n = x_n$ and $v_n = 0$ for the pilot/guard indices n. Next, we compute the LMMSE symbol estimates $\widehat{\boldsymbol{x}}$ and posterior symbol variance vector $\boldsymbol{\nu}^x$ as

$$\widehat{\boldsymbol{x}} = \boldsymbol{\mu} + \boldsymbol{G}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\mu}) \tag{96}$$

$$\boldsymbol{\nu}^x = \boldsymbol{v} - \mathrm{diag}(\boldsymbol{G}\boldsymbol{A}\mathrm{Diag}(\boldsymbol{v})), \tag{97}$$

where

$$\boldsymbol{A} \triangleq (1-\eta)(\boldsymbol{I}_K\otimes\widehat{\boldsymbol{H}}) \tag{98}$$

$$\boldsymbol{G} \triangleq \mathrm{Diag}(\boldsymbol{v})\boldsymbol{A}^{\mathsf{H}}\big(\boldsymbol{A}\mathrm{Diag}(\boldsymbol{v})\boldsymbol{A}^{\mathsf{H}} + \sigma_w^2\boldsymbol{I}\big)^{-1}. \tag{99}$$

We then convert the posterior mean and variance $\widehat{\boldsymbol{x}}$ and $\boldsymbol{\nu}^x$ to extrinsic quantities by solving for the $\widehat{q}_n$ and ν_n^{q} that yield $1/\nu_n^{\mathrm{x}} = 1/\nu_n^{\mathrm{q}} + 1/v_n$ and $\widehat{x}_n/\nu_n^{\mathrm{x}} = \widehat{q}_n/\nu^{\mathrm{q}} + \mu_n/v_n$, which is accomplished by

$$\nu_n^{\mathrm{q}} = \frac{v_n\nu_n^{\mathrm{x}}}{v_n - \nu_n^{\mathrm{x}}} \tag{100}$$

$$\widehat{q}_n = \frac{\widehat{x}_n v_n - \mu_n\nu_n^{\mathrm{x}}}{v_n - \nu_n^{\mathrm{x}}}. \tag{101}$$

Finally we convert the extrinsic means and variances $\widehat{q}_n$ and ν_n^{q} into extrinsic coded-bit probabilities using (54) and (59), and pass them to the decoder. The decoder treats them as coded-bit priors, computes coded-bit posteriors, and passes the extrinsic information back to the LMMSE equalizer to begin the next turbo iteration.

As a result of the matrix inverse in (99), the LMMSE scheme (96)–(99) incurs a complexity of $O(KM^3)$ multiplies per block of KM symbols, or $O(M^2)$ multiplies per symbol. Compared to the $O(\log M)$ per-symbol per-iteration complexity of PBiGAMP, this is not favorable with regards to the scaling versus M. However, if in (99) we approximate the vector $\boldsymbol{v}$ by its average value, then the per-symbol complexity could be reduced to $O(\log M)$, since $\widehat{\boldsymbol{H}}$ is circulant and thus amenable to fast convolution. In particular, this LMMSE approximation would use $4K+1$ FFTs per symbol block (i.e., 1 to compute the eigenvalues of $\widehat{\boldsymbol{H}}$, $2K$ for the multiplication by $\boldsymbol{A}$ in (96), and $2K$ for the multiplication by $\boldsymbol{G}$ in (96)). Since PBiGAMP uses $4K+2$ FFTs, its per-iteration complexity would be only slightly higher. Of course, PBiGAMP performs several iterations. Still, we show in Section V-D that the total computational complexity of PBiGAMP is only a bit higher than the fast LMMSE scheme, in part because it requires fewer turbo iterations on average.

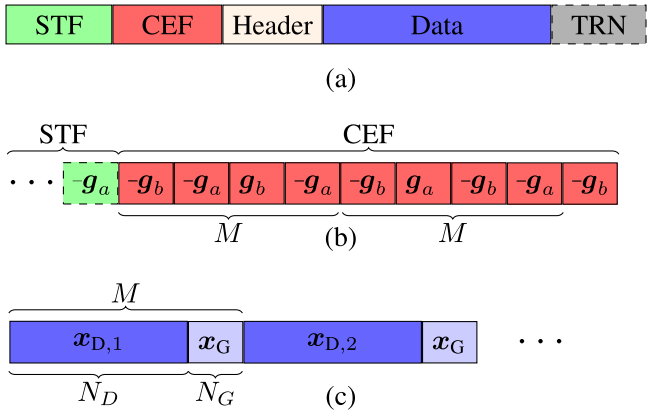


Fig. 4. (a) SC packet structure in the IEEE 802.11ad standard, including the Short Training Field (STF), Channel Estimation Field (CEF), Header field, Data field, and optional Training (TRN) field for beamforming; (b) inner structure of the CEF, constructed from length-128 Golay complementary sequences $\{\boldsymbol{g}_a, \boldsymbol{g}_b\}$; and (c) inner structure of the Data block, composed of data sequences $\{\boldsymbol{x}_{\mathrm{D},1}, \boldsymbol{x}_{\mathrm{D},2}\}$ and guard intervals $\boldsymbol{x}_{\mathrm{G}}$.

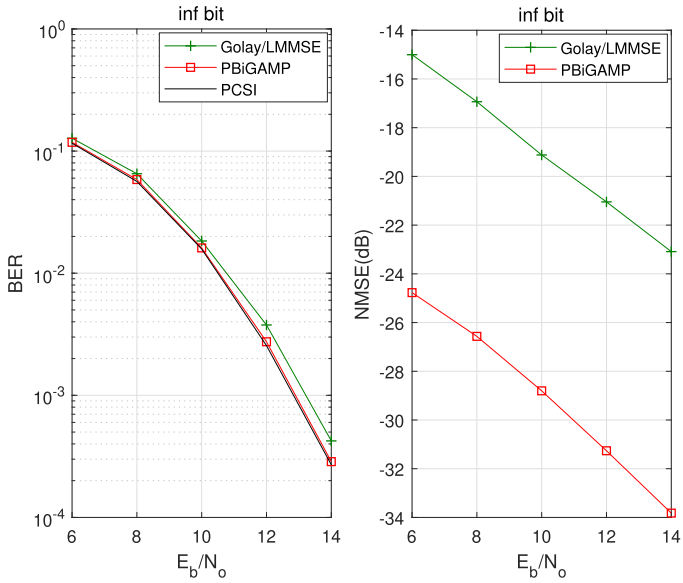


Fig. 5. BER and channel NMSE versus E_b/N_o in dB for 16-QAM with ∞-bit ADC under 60 GHz WLAN "conference room" channel.

V. Numerical Results

We now present numerical results comparing the proposed PBiGAMP method with the benchmarks discussed in Section IV. As a reference, we also consider the performance of PBiGAMP with perfect channel-state information (PCSI). In this latter case, PBiGAMP reduces to GAMP.

A. Setup

Unless otherwise noted, our numerical experiments are based on the following setup, which is compatible with the 802.11ad standard [2]. Recalling the SC block-transmission model from Section II-A, $N_{\mathrm{b}} = 3584$ information bits were coded at rate $R = 1/2$ by an irregular low-density parity-check (LDPC) code with average column weight 3, as specified by [2]. The 7168 coded bits were then Gray-mapped to 1792 16-QAM symbols (i.e., $A = 4$). The data symbols were then partitioned into $K_{\mathrm{D}} = 4$ blocks of $N_{\mathrm{D}} = 448$ symbols, resulting in $\{\boldsymbol{x}_{\mathrm{D}}[k]\}_{k=1}^{4}$. Each data-symbol sequence $\boldsymbol{x}_{\mathrm{D}}[k]$ was merged with an $N_{\mathrm{G}} = 64$-length guard sequence $\boldsymbol{x}_{\mathrm{G}}$, resulting in a $M = 512$-length data-guard sequence. The set was then merged with $K_{\mathrm{P}} = 2$ blocks of $M = 512$ pilot symbols, as shown in Figs. 1 and 4.

The 802.11ad standard specifies the use of Golay sequences [57] for constructing both $\boldsymbol{x}_{\mathrm{P}}$ and $\boldsymbol{x}_{\mathrm{G}}$. In particular, the pilot $\boldsymbol{x}_{\mathrm{P}}$ is constructed using the Golay complementary sequences $\{\boldsymbol{g}_a, \boldsymbol{g}_b\}$ as shown in Fig. 4(b), where both $\boldsymbol{g}_a$ and $\boldsymbol{g}_b$ have length $M/4 = 128$, and the guard $\boldsymbol{x}_{\mathrm{G}}$ is generated by an $N_{\mathrm{G}} = 64$-length Golay sequence. A correlation-based channel-estimation scheme that exploits the perfect aperiodic correlation property of Golay sequences is described in [56, Section 7.3.3.1]. We used that scheme for the benchmark described in Section IV-B, as well as to initialize the proposed PBiGAMP approach.

For the channel, we adopted the 60 GHz WLAN model described in Section II-C, whose Matlab implementation was obtained from [47]. We used the "conference room" scenario at baud rate 1.76 GHz with default parameter settings. Interestingly, the delay spread of this channel exceeds the guard length ($N_{\mathrm{G}} = 64$), implying some amount of inter-block interference (IBI). However, the PDP in Fig. 2(b) suggests that the IBI power is relatively small.

In the experiments below, one should remember that E_b/N_o values correspond to post-beamforming SNRs, which include the gain of beamforming at both the transmitter and receiver. In multi-antenna systems, the pre-beamforming SNRs are much lower.

B. BER and NMSE Performance With $\pi/2$-16-QAM

Figures 5–8 show the bit error rate (BER) and the channel-estimation normalized MSE (NMSE) versus E_b/N_o for ADCs with ∞-bit, 4-bit, 3-bit, or 2-bit precision. With an ∞-bit ADC (i.e., no quantization), PBiGAMP achieves a BER that is nearly indistinguishable from the PCSI bound, while Golay/LMMSE is 0.4 dB worse in BER and 10 dB worse in NMSE. With a 4-bit ADC the results are similar: PBiGAMP and PBiGAMP-Bussgang achieve BERs nearly indistinguishable from the PCSI bound (which has degraded 0.25 dB from the ∞-bit case), while Golay/LMMSE is 0.5 dB worse in BER and 10 dB worse in NMSE. With a 3-bit ADC, PBiGAMP's BER is still nearly indistinguishable from the PCSI bound (which has degraded 0.8 dB from the ∞-bit case), while that of PBiGAMP-Bussgang is 0.7 dB worse and Golay/LMMSE is 0.9 dB worse in BER and 10 dB worse in NMSE. With a 2-bit ADC, PBiGAMP's BER is still nearly indistinguishable from the PCSI bound (which has degraded 3.2 dB from the ∞-bit case), but the PBiGAMP-Bussgang and Golay/LMMSE BER traces show a large gap from the PCSI bound at high E_b/N_o. The 2-bit NMSE traces are non-monotonic as a result of the "stochastic resonance" phenomenon [8], [24], referring to the phenomemon where noise improves the performance of a nonlinear system [58].

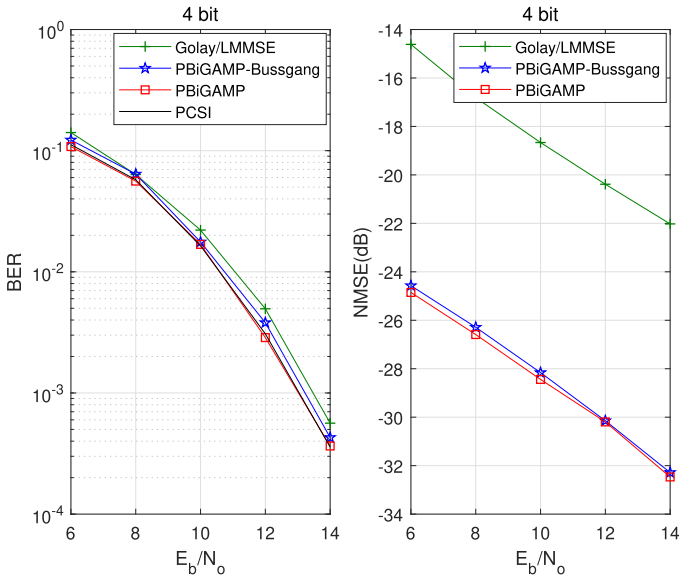


Fig. 6. BER and channel NMSE versus E_b/N_o in dB for 16-QAM with 4-bit ADC under 60 GHz WLAN "conference room" channel.

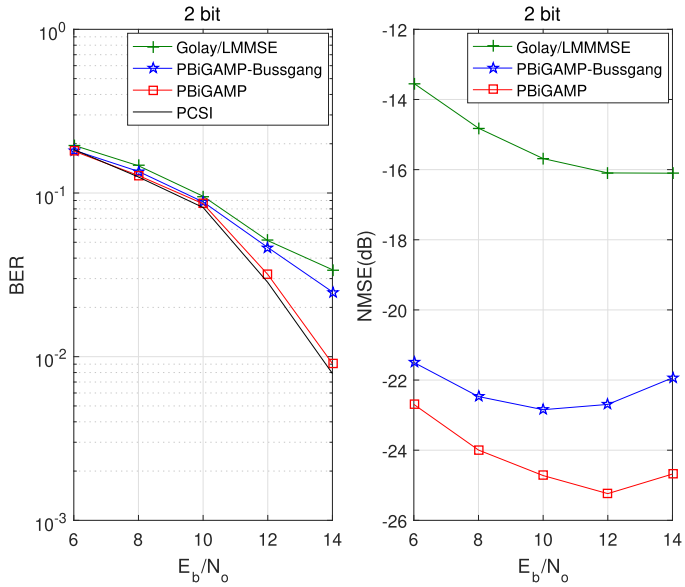


Fig. 8. BER and channel NMSE versus E_b/N_o in dB for 16-QAM with 2-bit ADC under 60 GHz WLAN "conference room" channel.

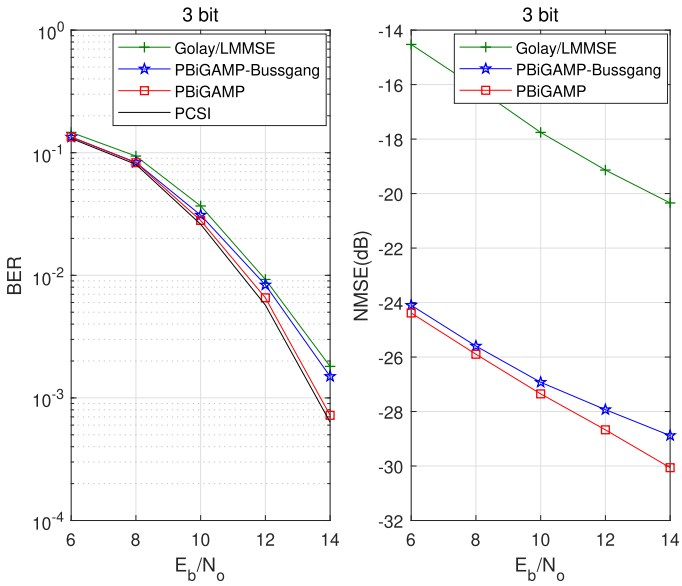


Fig. 7. BER and channel NMSE versus E_b/N_o in dB for 16-QAM with 3-bit ADC under 60 GHz WLAN "conference room" channel.

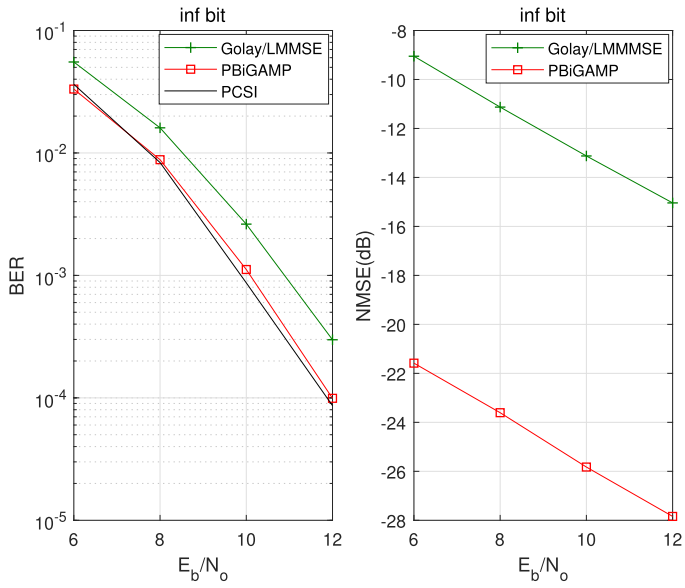


Fig. 9. BER and channel NMSE versus E_b/N_o in dB for $\pi/2$-BPSK with ∞-bit ADC under 60 GHz WLAN "conference room" channel.

C. BER and NMSE Performance With $\pi/2$-BPSK

In our experiments with 1-bit ADC, we found that none of the schemes under test were able to reliably decode the 16-QAM transmission described in Section V-B. We now show that 1-bit reception is feasible for $\pi/2$-BPSK transmissions, which is a mandatory mode of the 802.11ad standard [2]. For this, we coded $N_{\mathrm{b}} = 896$ information bits as before (i.e., at rate $R = 1/2$ using an irregular LDPC code with average column weight 3). The 1792 coded bits were then randomly interleaved and Gray-mapped to $N_{\mathrm{D}} = 1792$ symbols using $\frac{\pi}{2}$-BPSK (which rotates a standard BPSK transmission by $\pi/2$ radians each baud interval for improved PAPR). All other settings were the same as described earlier.

Figs. 9–12 show the bit error rate (BER) and the channel-estimation normalized MSE (NMSE) versus E_b/N_o for ADCs with ∞-bit, 3-bit, 2-bit, and 1-bit precision, respectively. With an ∞-bit ADC (i.e., no quantization), PBiGAMP achieves a BER that is nearly indistinguishable from the PCSI bound, while Golay/LMMSE is 0.9 dB worse in BER and 13 dB worse in NMSE. With a 3-bit ADC the results are similar: PBiGAMP and PBiGAMP-Bussgang achieve BERs nearly indistinguishable from the PCSI bound (which has degraded 0.3 dB from

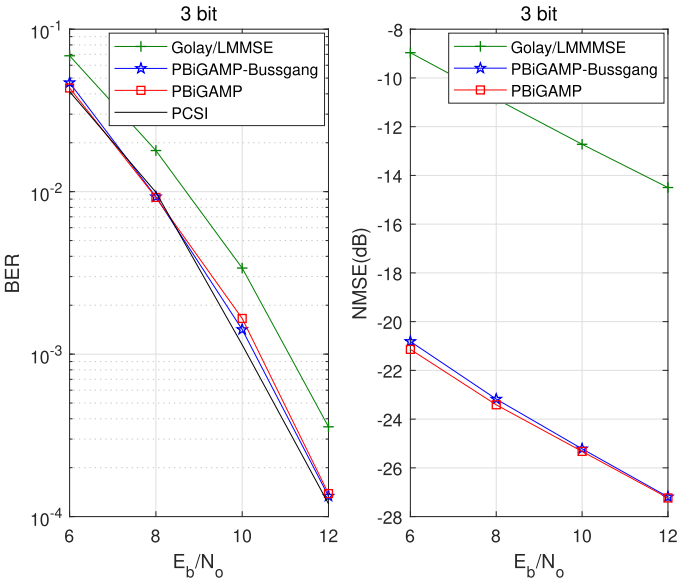


Fig. 10. BER and channel NMSE versus E_b/N_o in dB for $\pi/2$-BPSK with 3-bit ADC under 60 GHz WLAN "conference room" channel.

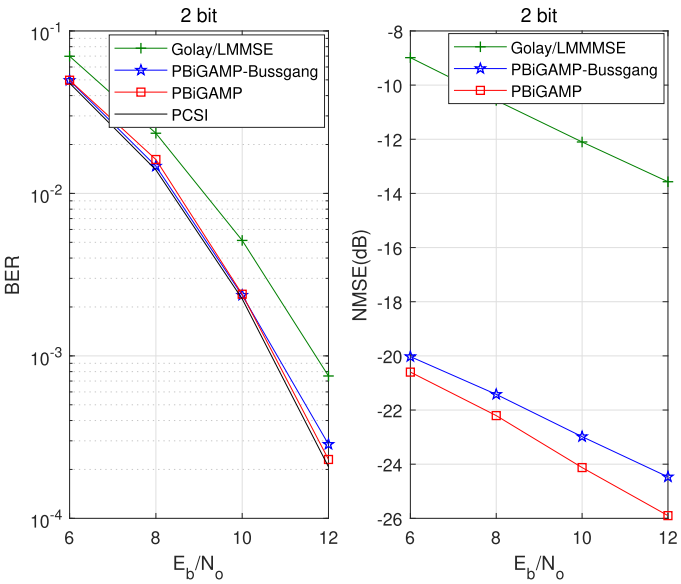


Fig. 11. BER and channel NMSE versus E_b/N_o in dB for $\pi/2$-BPSK with 2-bit ADC under 60 GHz WLAN "conference room" channel.

the ∞-bit case), while Golay/LMMSE is 0.9 dB worse in BER and 13 dB worse in NMSE. With a 2-bit ADC, the BERs of PBiGAMP and PBiGAMP-Bussgang are nearly indistinguishable from the PCSI bound (which has degraded 0.6 dB from the ∞-bit case), while Golay/LMMSE is 1 dB worse in BER and 13 dB worse in NMSE. With a 1-bit ADC, PBiGAMP's BER is still nearly indistinguishable from the PCSI bound (which has degraded 2.2 dB from the ∞-bit case), but the PBiGAMP-Bussgang and Golay/LMMSE BER traces show a large gap from the PCSI bound at high E_b/N_o. The 1-bit NMSE traces are non-monotonic as a result of the "stochastic resonance" phenomenon [54].

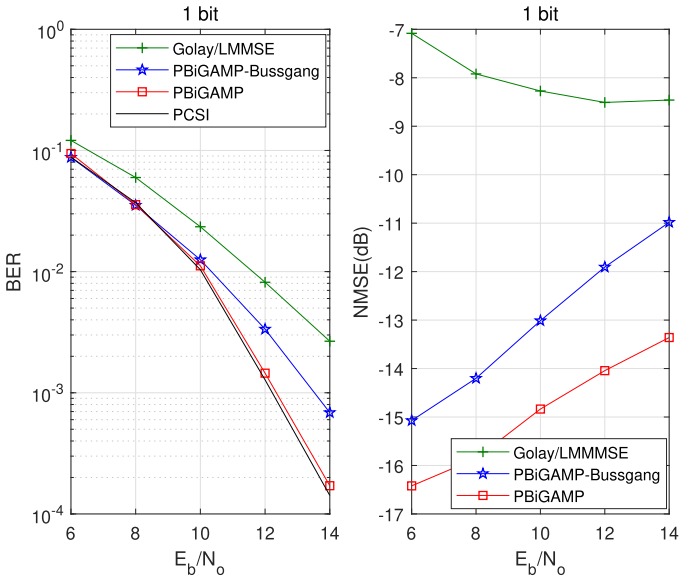


Fig. 12. BER and channel NMSE versus E_b/N_o in dB for $\pi/2$-BPSK with 1-bit ADC under 60 GHz WLAN "conference room" channel.

D. BER Versus Runtime With 16-QAM

To assess the computational complexity of PBiGAMP relative to the benchmark methods, we now present the results of runtime experiments in Matlab on a 3.3 GHz CPU.[4] The algorithms under test were PBiGAMP, Bussgang-linearized PBiGAMP, the exact Golay/LMMSE scheme (96)–(99), and the fast approximate Golay/LMMSE scheme described at the end of Section IV-B. PBiGAMP was terminated at the smallest iteration $t \geq 7$ at which $\sum_{m,k} |\widehat{x}_{mk}[t+1] - \widehat{x}_{mk}[t]|^2 < 0.01 \sum_{m,k} |\widehat{x}_{mk}[t+1]|^2$.

Figs. 13 and 14 plot BER versus average runtime for 16-QAM modulation and $E_b/N_o = 14$ dB at 2-bit and 3-bit quantization, respectively. The markers in each trace show the average BER and the average (cumulative) runtime at the end of each turbo iteration, indexed from 1 through 20. For each Monte-Carlo trial, a parity check was used to determine whether the BER was zero at the beginning of each turbo iteration and, if so, the equalization and decoding operations in that iteration were skipped. Thus, the *average* runtime contribution of the ith turbo iteration decrease with the iteration index i, because it is more likely that the BER equals zero in later turbo iterations.

Fig. 13 shows that, with 2-bit quantization, the fastest output comes from Golay/LMMSE-Fast after a single turbo iteration. However, the corresponding BER is relatively poor. At 2 turbo iterations, PBiGAMP yields a much lower BER than all other schemes, while consuming the same runtime as only 3 turbo iterations of Golay/LMMSE-Fast. And PBiGAMP yields even lower BERs after > 2 turbo iterations. Overall, Fig. 13 shows that PBiGAMP's accuracy-complexity tradeoff is vastly superior to those of the other methods.

Fig. 14 shows similar behavior with 3-bit quantization. As before, Golay/LMMSE-Fast achieves the fastest decoding, but

[4]The runtimes would be much faster in an ASIC or FPGA implementation.

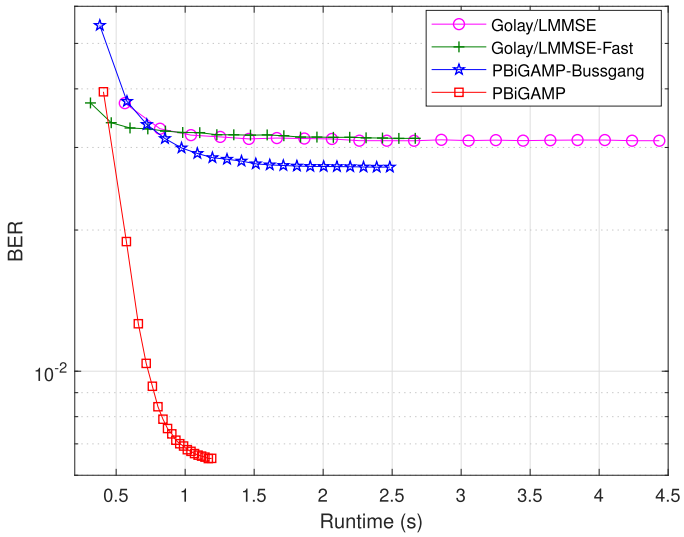


Fig. 13. BER versus average runtime for several algorithms with 16-QAM modulation and 2-bit quantization at $E_b/N_o = 14$ dB.

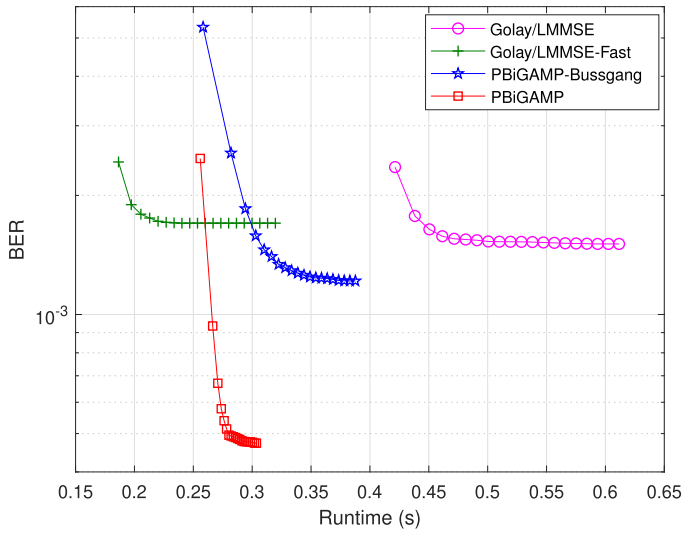


Fig. 14. BER versus average runtime for several algorithms with 16-QAM modulation and 3-bit quantization at $E_b/N_o = 14$ dB.

its BER is relatively poor. After only 2 turbo iterations, the BER of PBiGAMP surpasses the BERs achieved by all other methods. And the time it takes for PBiGAMP to complete 2 turbo iterations is only about 40% more than the time it takes for Golay/LMMSE-Fast to complete 2 turbo iterations. So, PBiGAMP gives a significant improvement in BER for a modest increase in complexity.

Several other observations can be made from Figs. 13–14. First the fast/approximate LMMSE scheme is much faster than the exact LMMSE scheme, although it yields slightly worse BER. Both behaviors are expected. Second, lower BER translates to faster average runtime per turbo iteration, because fewer turbo iterations need to be performed. So, more accurate equalization leads to improvements in runtime.

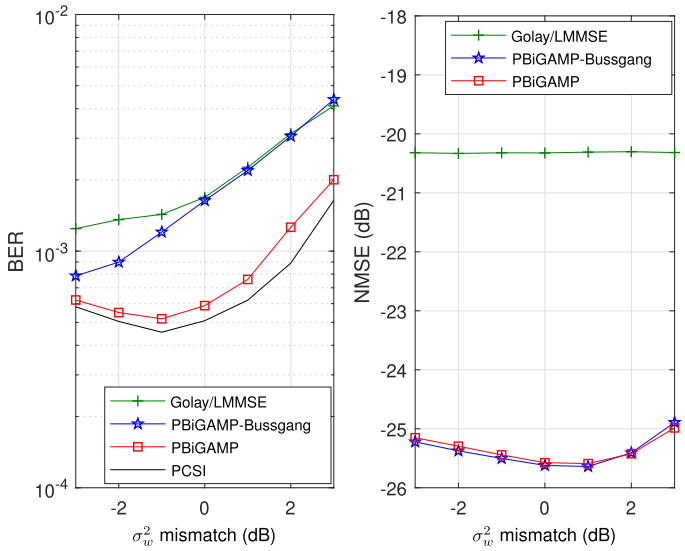


Fig. 15. BER and channel NMSE versus noise-variance mismatch in dB for 16-QAM with 3-bit quantization under the 60 GHz WLAN "conference room" channel at $E_b/N_o = 14$ dB.

E. Robustness to Noise-Variance Mismatch

Recall that all methods under test take the noise variance σ_w^2 as an imput. We now examine robustness to mismatch between the assumed and true values of σ_w^2.

Fig. 15 shows the BER and channel-estimation NMSE versus σ_w^2-mismatch in dB for 16-QAM with 3-bit ADC quantization at $E_b/N_o = 14$ dB. The figure shows that, as the assumed value of σ_w^2 grows larger than the true σ_w^2 (i.e., the mismatch in dB grows positive), the BERs of all methods degrade at a similar rate. However, as the assumed value of σ_w^2 grows smaller than the true σ_w^2 (i.e., the mismatch in dB grows negative), the BERs of all methods slightly improve before finally degrading. Fig. 15 also shows that PBiGAMP's channel estimation NMSE slightly degrades in the presence of noise-variance mismatch, while that of the Golay/LMMSE scheme remains relatively constant (but far worse than the value achieved by PBiGAMP).

Importantly, the BER of PBiGAMP closely tracks that of the perfect-CSI benchmark over the entire range of mismatch. This is the best possible outcome among schemes that take the noise variance σ_w^2 as an input parameter. Of course, it would be better to learn σ_w^2 from $\boldsymbol{y}$ rather than trust the supplied value of σ_w^2. As discussed in footnote 1, while extending PBiGAMP to learn σ_w^2 should not be difficult, we leave it for future work.

VI. Conclusions

In this paper we proposed a fast and near-optimal approach to joint channel-estimation, equalization, and decoding of coded SC transmissions over frequency-selective channels with few-bit ADCs. Our approach leverages the PBiGAMP algorithm to reduce the implementation complexity of joint channel estimation and symbol decoding to that of a few FFTs per iteration. Furthermore, it learns and exploits sparsity in the channel impulse response. Our work is motivated by millimeter-wave systems

with bandwidths on the order of Gsamples/sec, where few-bit ADCs, SC transmissions, and fast processing all lead to significant reductions in power consumption and implementation cost. We demonstrated our approach using signals and channels generated according to the IEEE 802.11ad wireless LAN standard, in the case that the receiver uses analog beamforming and a single ADC. Our experiments showed that the proposed approach yields BER almost indistinguishable from the known-channel oracle for ADCs with as few as 2-bit precision when recovering coded 16-QAM transmissions, and for ADCs with as few as 1-bit precision when recovering coded BPSK transmissions. Although it should be possible to recover coded QPSK transmissions with 1-bit ADCs, none of the schemes considered in this paper were able to do reliably with the 802.11ad codes and 802.11ad channels, and thus further work in this direction is warranted. As future work, it would also be interesting to extend our method to learn the noise variance σ_w^2 and to work with multiple few-bit ADCs, as in digital or hybrid beamforming systems.

REFERENCES

[1] R. W. Heath, N. González-Preclic, S. Rangan, W. Rho, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[2] *IEEE 802.11ad Standard Draft D0.1*, IEEE 802.11ad, 2012. [Online]. Available: www.ieee802.org/11/Reports/tgad_update.htm

[3] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[4] T. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[5] B. Murmann, "Energy limits in A/D converters," in *IEEE Faible Tension Faible Consommation*. Piscataway, NJ, USA: IEEE, Jun. 2013, pp. 1–4.

[6] B. Murmann, "ADC performance survey 1997–2018," 2018. [Online]. Available: http://www.stanford.edu/ murmann/adcsurvey.html

[7] A. Mezghani and J. A. Nossek, "Efficient reconstruction of sparse vectors from quantized observations," in *Proc. Int. ITG Workshop Smart Antennas*, 2012, pp. 193–200.

[8] J. Mo, P. Schniter, N. González-Prelcic, and R. W. Heath Jr., "Channel estimation in millimeter wave MIMO systems with one-bit quantization," in *Proc. Asilomar Conf. Signals Syst. Comput.*, Nov. 2014, pp. 957–961.

[9] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "Throughput analysis of massive MIMO uplink with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 4038–4051, Jun. 2017.

[10] A. Mezghani and A. L. Swindlehurst, "Blind estimation of sparse broadband massive MIMO channels with ideal and one-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2972–2983, Jun. 2018.

[11] Z. Zhou, X. Chen, D. Guo, and M. L. Honig, "Sparse channel estimation for massive MIMO with 1-bit feedback per dimension," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2017, pp. 1–6.

[12] A. Mezghani, M. S. Khoufi, and J. A. Nossek, "A modified MMSE receiver for quantized MIMO systems," in *Proc. Int. ITG Workshop Smart Antennas*, 2007, pp. 1–5.

[13] A. Mezghani, M.-S. Khoufi, and J. Nossek, "Spatial MIMO decision feedback equalizer operating on quantized data," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2008, pp. 2893–2896.

[14] A. Mezghani and J. Nossek, "Belief propagation based MIMO detection operating on quantized channel output," in *Proc. IEEE Int. Symp. Inf. Theory*, 2010, pp. 2113–2117.

[15] S. C. Wang, Y. Z. Li, and J. Wang, "Multiuser detection in massive spatial modulation MIMO with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2156–2168, Apr. 2015.

[16] Y. Xiong, N. Wei, and Z. Zhang, "A low-complexity iterative GAMP-based detection for massive MIMO with low-resolution ADCs," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2017, pp. 1–6.

[17] C. K. Wen, C. J. Wang, S. Jin, K. K. Wong, and P. Ting, "Bayes-optimal joint channel-and-data estimation for massive MIMO with low-precision ADCs," *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2541–2556, May 2016.

[18] F. Steiner, A. Mezghani, A. L. Swindlehurst, J. A. Nossek, and W. Utschick, "Turbo-like joint data-and-channel estimation in quantized massive MIMO systems," in *Proc. Int. ITG Workshop Smart Antennas*, 2016, pp. 1–5.

[19] O. Dabeer and U. Madhow, "Channel estimation with low-precision analog-to-digital conversion," in *Proc. IEEE Int. Conf. Commun.*, 2010, pp. 1–6.

[20] G. Zeitler, G. Kramer, and A. Singer, "Bayesian parameter estimation using single-bit dithered quantization," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2713–2726, Jun. 2012.

[21] A. Mezghani, F. Antreich, and J. Nossek, "Multiple parameter estimation with quantized channel output," in *Proc. Int. ITG Workshop Smart Antennas*, 2010, pp. 143–150.

[22] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. Swindlehurst, and L. Liu, "Channel estimation and performance analysis of one-bit massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4078–4089, Aug. 2017.

[23] T. Lok and V.-W. Wei, "Channel estimation with quantized observations," in *Proc. IEEE Int. Symp. Inf. Theory.*, Aug. 1998, p. 333.

[24] J. Mo, P. Schniter, and R. W. Heath Jr., "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1141–1154, Mar. 2018.

[25] S. C. Wang, Y. Z. Li, and J. Wang, "Multiuser detection for uplink large-scale MIMO under one-bit quantization," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2014, pp. 4460–4465.

[26] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency domain equalization for single-carrier broadband wireless systems," *IEEE Commun. Mag.*, vol. 40, no. 4, pp. 58–66, Apr. 2002.

[27] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Mag.*, vol. 28, no. 5, pp. 5–14, May 1990.

[28] A. Maltsev, R. Maslennikov, A. Sevastyanov, A. Khoryaev, and A. Lomayev, "Experimental investigations of 60 GHz WLAN systems in office environment," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1488–1499, Oct. 2009.

[29] T. Rappaport, F. Gutierrez, E. Ben-Dor, J. Murdock, Y. Qiao, and J. Tamir, "Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications," *IEEE Trans. Antennas Propag.*, vol. 61, no. 4, pp. 1850–1859, Apr. 2013.

[30] M. Akdeniz, Y. Liu, S. Sun, S. Rangan, T. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[31] 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.

[32] P. Schniter and A. Sayeed, "Channel estimation and precoder design for millimeter-wave communications: The sparse way," in *Proc. Asilomar Conf. Signals Syst. Comput.*, 2014, pp. 273–277.

[33] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inf. Thy.*, Aug. 2011, pp. 2168–2172, (full version at arXiv:1010.5141).

[34] C. Cao, H. Li, and Z. Hu, "An AMP based decoder for massive MU-MIMO-OFDM with low-resolution ADCs," in *Proc. Int. Conf. Comput., Netw., Commun.*, 2017, pp. 449–453.

[35] C. Douillard, M. Jezequel, C. Berrou, A. Picart, P. Didier, and A. Glavieux, "Iterative correction of intersymbol interference: Turbo equalization," *Eur. Trans. Telecommun.*, vol. 6, pp. 507–511, Sep./Oct. 1995.

[36] R. Koetter, A. C. Singer, and M. Tüchler, "Turbo equalization," *IEEE Signal Process. Mag.*, vol. 21, no. 1, pp. 67–80, Jan. 2004.

[37] J. T. Parker and P. Schniter, "Parametric bilinear generalized approximate message passing," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 4, pp. 795–808, Jun. 2016.

[38] P. Schniter, "Turbo reconstruction of structured sparse signals," in *Proc. Conf. Inf. Sci. Syst.*, Princeton, NJ, USA, Mar. 2010, pp. 1–6.

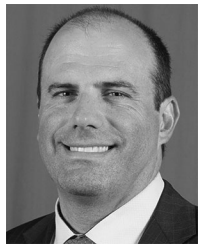
[39] P. Schniter, "A message-passing receiver for BICM-OFDM over unknown clustered-sparse channels," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 8, pp. 1462–1474, Dec. 2011.

[40] 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.

[41] P. Sun, Z. Wang, R. W. Heath Jr., and P. Schniter, "Joint channel-estimation/decoding with frequency-selective channels and few-bit ADCs," in *Proc. Asilomar Conf. Signals Syst. Comput.*, 2017, pp. 1824–1828.
[42] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath Jr.,"Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
[43] H. Yan, S. Ramesh, T. Gallagher, C. Ling, and D. Cabric, "Performance, power, and area design trade-offs in millimeter-wave transmitter beamforming architectures," arXiv:1807.07201, 2018.
[44] A. Maltsev *et al.*, "Channel models for 60 GHz WLAN systems," IEEE, Piscataway, NJ, USA, Tech. Rep. 802.11-09/0334r8, 2010.
[45] J. Ziniel, P. Schniter, and P. Sederberg, "Binary classification and feature selection via generalized approximate message passing," *IEEE Trans. Signal Process.*, vol. 63, no. 8, pp. 2020–2032, Apr. 2015.
[46] J. Max, "Quantizing for minimum distortion," *IRE Trans. Inf. Theory*, vol. 6, no. 1, pp. 7–12, 1960.
[47] R. Maslennikov and A. Lomayev, "Implementation of 60 GHz WLAN channel model," IEEE, Piscataway, NJ, USA, Tech. Rep. 802.11-10/0854r3, 2010.
[48] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2007.
[49] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.
[50] G. F. Cooper, "The computational complexity of probabilistic inference using Bayesian belief networks," *Artif. Intell.*, vol. 42, pp. 393–405, 1990.
[51] T. Minka, "A family of approximate algorithms for Bayesian inference," Ph.D. dissertation, Dept. Comput. Sci. Eng., MIT, Cambridge, MA, USA, Jan. 2001.
[52] C. Schülke, P. Schniter, and L. Zdeborová, "Phase diagram of matrix compressed sensing," *Physical Rev. E*, vol. 94, no. 6, Dec. 2016, Art. no. 062 136.
[53] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.
[54] A. Mezghani and J. Nossek, "Capacity lower bound of MIMO channels with output quantization and correlated noise," in *Proc. IEEE Int. Symp. Inf. Theory.*, 2012, pp. 2113–2117.
[55] J. A. Bucklew and N. C. Gallagher Jr.,"Some properties of uniform step size quantizers," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 5, pp. 610–613, Sep. 1980.
[56] T. S. Rappaport, R. W. Heath Jr., R. C. Daniels, and J. N. Murock, *Millimeter Wave Wireless Communications*. London, U.K.: Pearson Education, 2014.
[57] M. Golay, "Complementary series," *IRE Trans. Inf. Theory*, vol. 7, no. 2, pp. 82–87, Apr. 1961.
[58] M. D. McDonnell, N. G. Stocks, C. E. M. Pearce, and D. Abbott, *Stochastic Resonance: From Suprathreshold Stochastic Resonance to Stochastic Signal Quantization*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

Peng Sun received the B.S. degree in communication engineering from Zhengzhou University, Zhengzhou, China, in 2012, and the Ph.D. degree in information and engineering from Zhengzhou University, in 2018. From 2015 to 2017, he was a visiting Ph.D. student with the Department of Electrical and Computer Engineering, The Ohio State University. He is currently a Postdoctoral Scholar with Zhengzhou University. His research interests include signal processing and communication theory.

Zhongyong Wang received the B.S. and M.S. degrees in automatic control from the Harbin Shipbuilding Engineering Institute, Harbin, China, in 1986 and 1988, respectively, and received the Ph.D. degree in automatic control theory and application from Xi'an Jiaotong University, Xi'an, China, in 1998. Since 1988, he has been with Zhengzhou University, Zhengzhou, China, as a Lecturer with the Department of Electronics. From 1999 to 2002, he was an Associate Professor, and in 2002, he was promoted to a Professor with the Department of Communication Engineering. His general fields of interest cover numerous aspects within embedded systems, signal processing, and communication theory.



Robert W. Heath, Jr. (S'96–M'01–SM'06–F'11) received the B.S. and M.S. degrees from the University of Virginia, Charlottesville, VA, USA, in 1996 and 1997, respectively, and the Ph.D. from Stanford University, Stanford, CA, USA, in 2002, all in electrical engineering. From 1998 to 2001, he was a Senior Member of the Technical Staff then a Senior Consultant with Iospan Wireless Inc, San Jose, CA, USA, where he worked on the design and implementation of the physical and link layers of the first commercial MIMO-OFDM communication system. Since January 2002, he has been with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX, USA, where he is a Cullen Trust for Higher Education Endowed Professor, and is a Member of the Wireless Networking and Communications Group. He is also the President and CEO of MIMO Wireless Inc., Houston, TX, USA. He authored *Introduction to Wireless Digital Communication* (Prentice Hall, 2017) and *Digital Wireless Communication: Physical Layer Exploration Lab Using the NI USRP* (National Technology and Science Press, 2012), and co-authored *Millimeter Wave Wireless Communications* (Prentice Hall, 2014).

Dr. Heath has been a co-author of 16 award winning conference and journal papers including the 2010 and 2013 EURASIP Journal on Wireless Communications and Networking best paper awards, the 2012 Signal Processing Magazine best paper award, a 2013 Signal Processing Society best paper award, 2014 EURASIP Journal on Advances in Signal Processing best paper award, the 2014 and 2017 Journal of Communications and Networks best paper awards, the 2016 IEEE Communications Society Fred W. Ellersick Prize, the 2016 IEEE Communications and Information Theory Societies Joint Paper Award, and the 2017 Marconi Prize Paper Award. He received the 2017 EURASIP Technical Achievement award and is co-recipient of the 2019 IEEE Kiyo Tomiyasu Award. He was a Distinguished Lecturer and member of the Board of Governors in the IEEE Signal Processing Society. In 2017, he was selected as a Fellow of the National Academy of Inventors. He is also a licensed Amateur Radio Operator, a Private Pilot, a registered Professional Engineer in Texas. He is currently Editor-in-Chief of IEEE SIGNAL PROCESSING MAGAZINE.

Philip Schniter (S'92–M'93–SM'05–F'14) received the B.S. and M.S. degrees in electrical engineering from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 1992 and 1993, respectively, and the Ph.D. degree in electrical engineering from Cornell University, Ithaca, NY, USA, in 2000.

From 1993 to 1996, he was employed by Tektronix Inc., Beaverton, OR, USA as a Systems Engineer. After receiving the Ph.D. degree, he joined the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH, USA, where he is currently a Professor. In 2008–2009, he was a Visiting Professor with Eurecom, Sophia Antipolis, France, and Supelec, Gif-sur-Yvette, France. In 2016–2017, he was a Visiting Professor with Duke University, Durham, NC, USA. His research interests include signal processing, wireless communications, and machine learning. In 2002, he was the recipient of the NSF CAREER Award, in 2016 the IEEE Signal Processing Society Best Paper Award, and in 2017 and 2018 the Qualcomm Faculty Award. He currently serves on the IEEE Sensor Array and Multichannel Technical Committee and the IEEE Computational Imaging Technical Committee.